The role of fibroblast growth factor 21 in mediating altered metabolism of the Deletor mouse model for mitochondrial myopathy

Iida-Marja Kleine

Master's Thesis

Master's Degree Programme in Translational Medicine
Faculty of Medicine
University of Helsinki
2018

Du spottest mein um meine Niedrigkeit,
Doch Armut tut dem Adel nichts zu Leid.
Paul Heyse, Italienisches Liederbuch
Set to music by Hugo Wolf

Minä olen pöljä ja sinä olet pöljä.
Anthony de Mello, suom. Vuokko Rissanen

Abstract

This study investigates the metabolic consequences of a biomarker for mitochondrial myopathies, using the mouse as a model organism. The studied biomarker is fibroblast growth factor 21 (FGF21), which is secreted in high amounts from the diseased muscle tissue. It is an endocrine hormone that regulates lipid metabolism, and in healthy individuals it is mainly secreted from the liver. I utilized skeletal muscle samples from mice that were either wild type or had a mitochondrial myopathy, both with or without a whole-body knockout of FGF21. I analysed a data set from a targeted metabolomic experiment conducted on the skeletal muscle samples. The experiment was performed by our collaborator Vidya Velagapudi. Additionally I measured protein and mRNA expression of selected enzymes from the muscle samples.

This study shows, that the cytokine FGF21 contributes to the disease progression of mitochondrial myopathy. The aspects of pathophysiology it regulates were all found to center on the metabolic pathway of one carbon (1C) metabolism. Serine de novo synthesis shuttles glucose carbons into 1C metabolism. The transsulfuration pathway produces glutathione using carbon units from the 1C pathway. The results of this study show, that FGF21 mediates the upregulation of alternative carbon donors in one carbon metabolism, especially serine biosynthesis, and the elevated utilisation of carbon units in the transsulfuration pathway. Not all of the metabolic changes characteristic of mitochondrial myopathy were affected by FGF21, e.g. the upregulation of acyl carnitines seen in mitochondrial myopathy was not affected by the knock-out of FGF21.

Keywords: mitochondrial disease, myopathy, FGF21, biomarker, adPEO, Twinkle, metabolic pathway, one carbon metabolism, targeted metabolomics, MetaboAnalyst

Contents

Abbreviations

adPEO	autosomal dominant progressive external ophthalmoplegia
AARE	amino acid response element
bp	base pairs
BSA	bovine serum albumin
COX	cytochrome oxidase
csv	comma separated values
cDNA	complementary DNA
DNA	deoxyribonucleic acid
dsDNA	double stand DNA
dsDNAse	double strand DNAse
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
GC	*gastrocnemius*
GSEA	gene set enrichment analysis
HRP	horseradish peroxidase
IOSCA	infantile onset spinocerebellar ataxia
KO	knock-out
ML	machine learning
MM	mitochondrial myopathy
mRNA	messenger RNA
MS	mass spectrometry
MSEA	metabolite set enrichment analysis
mtDNA	mitochondrial DNA
nDNA	nuclear DNA
NMR	nuclear magnetic resonance
OXPHOS	oxidative phosphorylation
ORA	over representation analysis
PCA	principal component analysis
PLS-DA	partial least squares discriminant analysis

PVDF	polyvinylidene fluoride
PCR	polymerase chain reaction
QF	*quadriceps femoris*
RNA	ribonucleic acid
RF	random forests
RT-qPCR	real time quantitative PCR
SDH	succinate dehydrogenase
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween 20
TCA	tricarboxylic acid cycle
TE	Tris-EDTA buffer
THF	tetrahydrofolate
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
VIP	variable importance in projection
WT	wild type

1 Introduction

Mitochondria are cellular organelles most often remembered as "the powerhouse of the cell". True to the stereotype, they do dutifully generate most of the ATP any eukaryote needs. A cancer biologist may often think of mitochondria as the organelles which regulate apoptosis, in the hopes of reinstating in cancer cells the ability to commit suicide. Nonetheless, upon making a closer acquaintance with mitochondria one learns that these hard-working old endosymbionts of ours are involved in a substantial amount of diverse cellular processes beyond oxidative phosphorylation or the storage and release of calcium ions (Nunnari & Suomalainen 2012). In integrating complex metabolic and signalling pathways, mitochondria are a profoundly essential and influential part of the cell (Nunnari & Suomalainen 2012, Zong et al. 2016).

The diversity of mitochondrial functions is mirrored in the diversity and symptomatic variability of mitochondrial diseases (Gorman et al. 2016). A substantial amount of intermediary metabolism takes place in mitochondria, and these metabolic pathways have been found to both be affected in mitochondrial diseases and to contribute to pathogenesis. Studying mitochondrial diseases leads not only towards better treatment options, but also sheds light on the complex biology of both mitochondria and the whole cell they inhabit and control.

The fibroblast growth factor 21 (FGF21) has been established as a biomarker for mtDNA maintenance and mitochondrial translation disorders, where it is found upregulated in the serum of patients (Lehtonen et al. 2016, Suomalainen et al. 2011a). FGF21 is termed a "fasting hormone", since it is known to be induced by fasting and it acts to mobilise fatty acids from white adipose tissue (Inagaki et al. 2007). Although usually originating from the liver (Nishimura et al. 2000), in these pathologies FGF21 is secreted by the diseased skeletal muscle.

The upregulation of FGF21 was first connected to inherited disease, when it was found to be induced in mitochondrial myopathy in the Deletor mouse (Tyynismaa et al. 2010). The Deletor is a mitochondrial disease model (Tyynismaa et al. 2005), which manifests with a late-onset mitochondrial myopathy (MM) primarily observed in the skeletal muscle *quadriceps femoris* (QF). Indeed, FGF21 was secreted directly from the affected muscle fibers. Another feature of the Deletor phenotype are widespread metabolic changes in its skeletal and heart muscle tissues (Nikkanen et al. 2016). It is of interest to discern the pathological role of FGF21 in the diseases that present with it, and this can be studied by inactivating FGF21 from Deletor mouse, by genetically knocking out the gene.

2 Aims

General aim:

Characterising the metabolic consequences of FGF21 in the mitochondrial myopathy of the mouse.

Specific aims:

1) Exploring methods of analysis and the means to interpret the results, and assessing their suitability for this study.

2) Establishing whether FGF21 act to progress the symptoms of the disease, or to protect the affected organism from them.

3) Describing the metabolic alterations caused by FGF21.

3 Review of literature

3.1 Mitochondrial diseases

Mitochondrial diseases are genetic diseases that harbour mutations in genes whose protein products localise in mitochondria. The disease course leads to an impairment of cellular energy metabolism, typically causing symptoms in organs that heavily utilise oxidative phosphorylation. Often the disease-causing mutation is found directly in one of the proteins associated with the mitochondrial respiratory complexes I-V, or in proteins involved in mtDNA maintenance or mitochondrial protein synthesis. Taken all together, these diseases are the most common form of inherited metabolic disease (Thorburn 2004) and their total prevalence is estimated to be 1 in 5000 (Ylikallio & Suomalainen 2012). No curative treatment for them exists (Suomalainen 2011b). A comprehensive overview on the topic can be found in the recent review by Gorman et al. (2016).

3.1.1 Complexity in the disease manifestation

Mitochondrial diseases are characterised by complexity in both disease manifestation and mechanism. The defectively mutated gene can reside in either nuclear DNA (nDNA) or mitochondrial DNA (mtDNA). Any inheritance pattern is possible, any organ system can be affected, and symptoms are heterogenous – even between patients with similar syndromes. If the mutation is in mtDNA, also the level of heteroplasmy affects the severity of disease manifestation. Heteroplasmy refers to the existence of both defective and intact mtDNA content in the cell.

Defects in the same gene can lead to diseases with very different symptoms. E.g. infantile onset spinocerebellar ataxia (IOSCA) and a form of autosomal dominant progressive external ophthalmoplegia (adPEO) both harbour a mutation in the mitochondrial helicase TWINKLE, but the former is a neurogenerative disorder and the latter a myopathy (muscle disease) (Koskinen et al. 1994, Suomalainen et al. 1992). The reason for such non-overlapping tissue manifestations in the case of Twinkle diseases is not known.

Mitochondria are an important site of metabolism within the cell (Nunnari & Suomalainen 2012). Their metabolic pathways are connected to the cytosol, e.g. through the folate cycle, which runs partly within the mitochondrial matrix and partly outside of the organelle (Ducker & Rabinowitz 2017). Because mitochondria house ample and interconnected metabolic and signalling pathways, these should be studied in the context of mitochondrial disease. Different cells in different tissues inevitably have different metabolomes. Should metabolite levels play a role in disease mechanism,

will their study lead to insights about the elusive tissue specificity of symptoms.

3.1.2 Progressive external ophthalmoplegia

Progressive external ophthalmoplegia (PEO) is an adult onset mitochondrial myopathy, and its most visible symptom is the weakening of the extraocular muscles, which results in drooping eyelids. These symptoms may occur alone, or the patients may develop ataxia, parkinsonism, premature menopause or depression - depending on the causative gene in question. PEO can be caused by a range of different mutations in different proteins, and it is characterised by multiple deletions in the mitochondrial genome in affected tissues. (Copeland 2008, Ylikallio & Suomalainen 2012)

One of the defective proteins in PEO is the TWINKLE helicase, which was first characterised and reported in Spelbrink et al. (2001), when it was found in conjunction with the disease. Since the helicase is encoded by nDNA and the impaired allele is dominant, PEO with a TWINKLE defect is autosomal dominant PEO (adPEO). The defect in TWINKLE leads to repeated mtDNA replisome stalling and dsDNA breaks (Goffart et al. 2009). The dsDNA repair mechanism then causes a deletion of DNA at every repaired breakpoint (Srivastava & Moraes 2005), so adPEO is a mtDNA maintenance disorder. The mtDNA deletions accumulate over the lifetime of the patient, eventually crossing a phenotypic threshold when symptoms appear.

3.1.3 The Deletor mouse

The Deletor mouse is a model organism for a late onset progressive mitochondrial myopathy, originally designed in Tyynismaa et al. (2005). It was generated by inserting a TWINKLE patient mutation (Spelbrink et al. 2001) into a homologous position in the mouse genome. As a mitochondrial disease model the Deletor is remarkable and useful because it is based on a patient mutation and because the phenotype is that of a late onset disease, which does not have many model organisms (Tyynismaa & Suomalainen 2009).

The model was reported to replicate a range of relevant histological, genetic, and biochemical features of PEO patients (Tyynismaa et al. 2005). These include COX-/SDH+ skeletal muscle fibres, multiple-sized deletion accumulation at specific sites in the mitochondrial genome in muscle and brain tissues, and a similar age of onset of the disease when related to the expected lifespan. The age of onset is 12 months for the Deletor, based on the appearance of histological signs in skeletal muscle. The mice have a normal lifespan and gross motor performance despite of the disease.

3.1.4 The amino acid starvation response

The gene expression of Deletor skeletal muscle was studied in Tyynismaa et al. (2010), and a strong induction of genes under control of amino acid response elements (AARE) was found, among these the fibroblast growth factor 21 (FGF21) and the enzyme methylenetetrahydrofolate dehydrogenase 2 (MTHFD2), which acts in the folate cycle. The expression of both these proteins increased with the progression of the disease, implying their possible involvement in the disease mechanism.

Members of the activating transcription factor (ATF) family bind AARE sites at promoter regions of genes important for cellular stress reactions, notably upon amino acid limitation. ATF4 is deemed the master regulator of the following signalling cascade (Kilberg et al. 2005). FGF21 is mainly regulated by ATF4, through the promoter sequences AARE1–3 (Maruyama et al. 2016). Because of the involvement of AARE:s found in the Deletor, the myopathy was originally described as a state of "pseudostarvation". However, current knowledge suggests that anabolic biosynthesis pathways are chronically induced.

3.2 The fibroblast growth factor 21

Fibroblast growth factor 21 (FGF21) belongs within the FGF superfamily into a subgroup which has endocrine action, together with FGF19 and FGF23. These FGFs share the feature that they have poor binding affinity to heparin, which makes transport through the circulatory system and their function as endocrine hormones possible. FGF21 signals via FGF receptors by simultaneously binding the co-receptor βKlotho (Kharitonenkov & DiMarchi 2017).

FGF21 was discovered in 2000 (Nishimura et al. 2000) and brought to wider attention in 2005 in a screening panel because of its ability to regulate glycemic control (Kharitonenkov et al. 2005). It stimulates lipolysis and increases insulin sensitivity, thereby promoting resistance to obesity (Kharitonenkov et al. 2005) and causing weight loss (Kharitonenkov et al. 2006). There has been a lot of enthusiastic research centered on the aim of using the protein to treat diabetes and obesity (Sonoda et al. 2017).

The main source of FGF21 in healthy individuals is the liver in both mice and humans (Nishimura et al. 2000). FGF21 crosses the blood-brain-barrier and recent results indicate, that both the liver and the brain regulate the release of the hormone from the liver (Potthoff 2016). However, in mitochondrial myopathies FGF21 is secreted from skeletal muscle into circulation in such a robust manner, that is has been established as a biomarker for these diseases (Suomalainen et al. 2011a). This is an important finding concerning the diagnosis of mitochondrial diseases, where a muscle

biopsy is standard procedure for diagnosis of muscle manifesting diseaes. Circulating FGF21 can be detected from a simple blood sample.

FGF21 has been found to mediate the adaptive fasting response through $PPAR\alpha$, promoting torpor and inducing lipolysis of white adipose tissue and ketogenesis in the liver in mice (Inagaki et al. 2007). As a hormone, FGF21 is still a fairly new object of study and its actions are not well understood. Even downright paradoxical observations are published, e.g. mild hypoglycaemia in both FGF21 KO and transgenic FGF21 mice, as pointed out in Kharitonenkov & DiMarchi (2017).

The intricacies of FGF21 signaling are nonetheless being looked into, and regulatory actions of the hormone obtain new explanations through a more detailed picture of its physiology. One interesting aspect of it became apparent when it was found upregulated in mitochondrial myopathy: through the actions of FGF21, the influence of mitochondrial signalling extends from within the confinements of a single cell to distant tissues in the whole organism.

3.3 Metabolomics

Metabolomics is a younger 'omics field and like the others, e.g. transcriptomics, it has been made possible by technological advances that permit experiments of larger scale and higher performance. Instead of tens or hundreds of individual biochemical assays, one single metabolomic experiment can be performed to obtain a snapshot of the metabolic state in a given biological sample. The necessary technological advances were made in the synthesis of isotope labelled compounds and in instrumentation for analytical chemistry, and in the development of computational tools with which to analyse the measurement output. (Johnson et al. 2016)

The measurements are usually performed using either mass spectrometry (MS) or nuclear magnetic resonance (NMR) instruments, often coupled to gas chromatography (GC) or liquid chromatography (LC) for additional separation of the different chemical species. GC is suitable for easily volatile metabolites. After the physical experiment, preprocessing of the raw data, and thereafter statistical and bioinformatic analysis is needed to achieve meaningful results. (Liu & Locasale 2017)

3.3.1 Targeted and untargeted approaches

The field is commonly divided into two, based on the approach and the phrasing of research questions. One large category is untargeted metabolomics, and the other is targeted metabolomics. An untargeted metabolomics experiment is done without *a priori* knowledge about which metabolites

will appear in the measurement. The aim is to identify as many metabolites in the sample as possible, analogous to shotgun proteomics, and the results are usually qualitative in nature. Identification of new metabolic compounds is possible in untargeted experiments, though analysis of mass spectra with complex analyte mixtures can be challenging. (Ebbels & Iorio 2011*a*)

In targeted metabolomics, the experiment has been optimised for measurement of a specific set of compounds. Because of this optimisation, detection of each individual metabolite is robust and the results are usually quantitative. The approach requires prior knowledge about the presence of the targeted metabolites in the sample. It can be used for validation of results from untargeted experiments, or else when robust readings for specific metabolites are sought after. Sometimes an experiment is termed "semi-untargeted", implying that the set of measured compounds has been selected with little bias from biological knowledge. (Liu & Locasale 2017)

3.3.2 Metabolomic experiments yield large sets of data

An untargeted metabolomic experiment produces a large amount of high-dimensional spectral data. The most time-comsuming part of the untargeted experimental workflow is the preprocessing of this instrumental data, incorporating operations such as retention time alignment, spectral binning, and peak picking and annotation (Gowda et al. 2014). The result of preprocessing is a list of identified metabolites and simply being able to identify them takes all the effort, as is apparent e.g. in the master's thesis by Kalogeropoulou (2011). After this, qualitative analysis ensues.

In targeted metabolomics the compound identification and quantitation is a routine and previously optimised step of the worflow. Nonetheless, also here the most time-consuming affair is data analysis. This time the data at hand is numerical, so statistical analysis and quantitative approaches to bioinformatics become applicable and necessary. Simple "eyeballing" of the data is only a start to what can be done in terms of analysis (Chadeau-Hyam et al. 2013). The analysis then very quickly yields an abundance of lists, plots, and tables as results, and navigating the possibilites for use of analytical tools and therefore keeping track of the obtained results is of importance in both targeted and untargeted metabolomics.

3.3.3 Working around multiple testing

When performing statistical tests on large sets of data, the type I statistical error of false positives becomes commonplace. When e.g. comparing the Deletor to WT in this study using a statistical test, we have approximately one hundred metabolites to analyse and should therefore expect to

find five false positives on the significance level of $\alpha = 5\%$. Performing many statistical tests in this fashion is termed multiple testing (Chadeau-Hyam et al. 2013, Mary-Huard & Robin 2011).

There are many adjustments for multiple testing, aiming to correct p-values so that the type I error rate decreases. The perhaps best known one is the Bonferroni adjustment, which is widely used but quite strict and conservative (Armitage et al. 2015). Another approach is to calculate false discovery rates (FDR) in order to estimate the rate of type I error and stay aware of the possibility of false positives. New methods of FDR estimation are constantly being developed.

Controlling for type I errors in metabolomics is difficult, because the approach to data analysis is largely explorative. There is no simple and clear predefined hypothesis to be tested, which could be used to plan for all of the steps of statistical analysis in advance so that the error rate could be minimised. Rather, the analysis seeks to find interesting features and generate hypotheses based on these. In this case, adjusting p-values is meaningless because of the complexity of the analysis (Bender & Lange 2001). Instead, new biological replicates, measurements, and studies, or complementary independent experiments are needed to validate any findings of a metabolomics study. Suitable experiments are e.g. measuring mRNA or protein expression levels of those enzymes, whose substrates have attracted attention in the metabolomic data.

3.4 Tools for metabolomic analysis

A bioinformatician should always have time to just "play with the data" (personal communication, Chengyu Liu). No single bioinformatic protocol exists, not even a very thorough one, that could be used to extract all possible information out of a set of biological data. However, there are specific tools which have been found generally useful. By testing these tools, the bioinformatician gets an impression of the data and preliminary results, which guide the selection of further tools and approaches for a meaningful analysis. Knowledge is needed on which tools, why, and how to use, and how to interpret the results.

As metabolomics is becoming more widely used as an experimental approach, researchers need to learn how to interpret the metabolomic data. To aid this, developers of bioinformatic tools have been building environments, in which a layman can perform advanced statistical analyses with guidance for using methods and interpreting results. Notable such environments are e.g. XCMS (Tautenhahn et al. 2012, Gowda et al. 2014), and MetaboAnalyst (Xia et al. 2009), the latter of which has been used in this study. XCSM is designed particularly for untargeted metabolomics, and MetaboAnalyst can be used for both targeted and untargeted data. Below, key points about

some common tools for metabolomic analysis are explored and explained, namely what a specific tool does and how it works.

3.4.1 Preprocessing and outlier removal

Preprocessing is a mechanical step in the analysis pipeline, preparing the data for subsequent analysis. Certain unit operations have to be performed on a given set of data in order to obtain reliable results from it. As mentioned, in untargeted metabolomics it is particularly effortful. When the results from measurements are obtained as concentration values in targeted metabolomics, only statistical preprocessing is needed before statistical analysis. This usually comprises normalization, transformation, or scaling, and possibly outlier removal before these (Xia & Wishart 2016).

Multivariate methods and e.g. log-scaling cannot cope with missing values (Xi et al. 2014), which is why these either have to be replaced, or the whole sample or feature with missing values removed from analysis. Because of this, removal of single outlier values should be avoided. Normalisation aims to remove or dampen unwanted noise in the data, thereby also treating possible outliers so as to lessen their distorting effect. If outliers arise and are removed from the analysed data, normalisation has to be redone.

It would be best if there always was an instrumental reason for outlier removal (personal communication, Sampsa Hautaniemi). In the strict sense, computational methods cannot reliably account for outliers even though many statistical tests for outlier removal exist. A practical observation of measurement error is the preferred justification for outlier removal. On the other hand, whole samples are often judged as outliers based on dimension reduction models, in which they may stand out from the rest of the sample, e.g. a single sample that shows up very far away from a cluster of the sample group on a PCA scores plot.

3.4.2 Visualisation of the data

Most of the analysis methods produce a visual result as an outcome, and such one should be sought after in all other cases, too. Visualisation helps the mind grasp the data. It is good to start exploring a new data set by generating different plots of it, e.g. using simple univariate statistics. These plots for data exploration will probably not end up in publications, but are instead a useful tool to familiarise oneself with the data. (Chang 2013)

A useful way to get an overview and see all the data at once is to generate a heatmap of the concen-

tration values. MetaboAnalyst offers hierarchical clustering (HC) in conjunction with heatmaps, so that samples and metabolites are organised based on the HC dendrogram. The distance measure for HCA can be chosen between the Pearson R, Spearman rank correlation, and Kendall rank correlation. When exploring the data, it is useful to not cluster the samples, so that trends within and between sample groups become evident in the heatmap, when samples from the same group are displayed next to each other (Xia & Wishart 2016). In an optimal case the HC algorithm would judge samples from the same group as similar and place them next to each other, but this oftentimes does not happen and hence not clustering the samples helps.

Box plots are an effective way of fitting both data values and descriptive statistics into one picture. The ranges of the two middle and two outer quartiles of data points are depicted by the box and its whiskers, respectively. The median is drawn as a horizontal line within the box, and outlier candidates are drawn as individual points outside the reach of the whiskers. Box plots can easily be drawn e.g. in R with a single function applied on values for one feature. Viewing the plots gives insight into patterns of the values, to see which are up or down, and into the spread of values, are they centered around the median or scattered more widely. (Chang 2013)

The more detailed positioning of values relative to the median or mean can be shown in dot plots, where each individual value is plotted alongside a horizontal line marking the mean/median. A variation of the box plot with an advantage similar to the dot plot is the violin plot, where the rectangular contours of the box are replaced by a curve showing the distribution of values. Volcano plots are useful for showing both magnitude and significance of changes in the data. Here, the $\log_2$ of fold change (FC) is plotted over the negative $\log_{10}$ of the p-value. Large and significant changes become apparent in the upper corners of the plot, along with their amount relative to all data points. (Chang 2013)

3.4.3 Multivariate methods for dimension reduction and classification

Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) are multivariate analytical methods, which aim to identify features of the data that most contribute to variation or separation (Worley & Powers 2013). They are designed to reduce the amount of dimensions of the data while preserving information on variance. PCA is an unsupervised and PLS-DA is a supervised method, that is, the latter algorithm uses the predefined class labels for calculation in addition to the data itself.

PCA gives an unbiased result and it is useful for a first look at what kind of separation can be found in the data. However, for PCA to be able to separate groups, the within-group variation has

to be less than between-group variation (Worley & Powers 2013), which is why PLS-DA is often used in addition or instead of PCA. In an ideal case PLS-DA could be used to verify and further parse the clustering and loadings first observed in PCA. PCA can be used for outlier detection. An outlier sample will be located far from other clustered samples on the scores plot (Xia & Wishart 2016).

The goodness of fit of a PLS model is given by R2 and Q2, where R2 gives the upper and Q2 the lower limit of how well the data is explained by the model and how well the model predicts new observations (Wold et al. 2001). In MetaboAnalyst these values are calculated in conjunction with the cross validation and they should be checked before interpretation of the results. Because PLS is a supervised method, it is prone to overfitting and finding separation in any set of data. An overfitted model cannot predict observations in new measurements. The problem of overfitting is minimised by choosing a low-dimensional model and not using excessive components for analysis (Ebbels & Iorio 2011*b*).

Both tools give scores and loadings plots as results. The scores are the weights of model components for each sample, and loadings express the weights of individual features (metabolites) in forming the principal component. The models are multidimensional vectors, and the scores plots are two or three dimensional projections of these vectors used for visualisation. Usually the first two components are plotted, though it is useful to explore the model by looking at plots of different component pairs, e.g. the second and third component. (Xia & Wishart 2016, Ebbels & Iorio 2011*b*)

Each score has respective loadings, and important metabolites are found by looking at the loadings plot specific for a chosen scores plot. Metabolites which are located in the same direction as a cluster of samples are important for the classification of those samples, e.g. a cluster with a high score on component 1 on the scores plot is classified based on metabolites high on the loadings 1 axis. This method of interpretation is used for both PCA and PLS-DA scores and loadings plots (Xia & Wishart 2016).

3.4.4 Machine learning in classification: Random forests

Random forests (RF) is a machine learning algorithm used for classification (Ressom et al. 2008). In it, a large amount of decision trees is built, each tree starting with a random feature. As the trees are then used for classification of samples from the data, they vote for the most popular class and assign it to the sample. At each node of a tree, features are sampled with the bootstrapping method. It can be mathematically shown that the trees always converge, which means that overfitting is not a problem of this method (Breiman 2001).

MetaboAnalyst offers RF for classification and for outlier sample detection. The algorithm reports an overall out of bag (OOB) error and a plot of class error rates over the count of trees. The OOB error is estimated based on data that is left out during tree construction (Breiman 2001). The plot should be used for checking, after how many generated trees the class error rates become stable and adjusting the amount of trees so that it exceeds that limit (Xia & Wishart 2016). The stable error rate means that the value stays constant over a growing number of trees.

3.4.5 Detecting significant changes

Simple and widely used statistical tests such as ANOVA, Students' t-test or the Wilcoxon Rank Sum test can be applied to metabolomic data while keeping in mind, that these do not automatically include adjustments for multiple testing. If testing is done on original unnormalised data, it is useful to investigate the normality of the data in order to determine if a test can be used that assumes a normal distribution (Mary-Huard & Robin 2011). In general, several different approaches to analysis yielding similar results solidifies the results. Results from different tests can be compared for their greater reliability.

Significance analysis of microarrays (SAM) is a statistical method that yields a list of significantly changed features as a result, and it can be used in MetaboAnalyst. It was originally developed for analysing microarray data (Tusher et al. 2001), but it has been adapted for use in metabolomics (Xia & Wishart 2016). The analysis is guided by specifying a delta value that controls the FDR. A larger delta gives a smaller FDR and a smaller amount of significantly changed metabolites.

3.4.6 Hunting for patterns

Often, patterns in the levels of metabolites in the studied groups begin to emerge in the data analysis. Using correlation coefficients, more metabolites with the same observed concentration pattern can be found. In MetaboAnalyst, correlation coefficients are utilised in correlation matrices and in the PatternHunter tool (Xia & Wishart 2016). In the correlation matrices Pearson R values calculated for all metabolite pairs are plotted on a colour scale. The metabolite rows can additionally be organised with hierarchical clustering in the same way as in the heatmaps. Distinct areas of high positive or negative correlation can then be looked at to see which metabolites form them.

The PatternHunter uses the same correlation coefficients as can be chosen for HCA. Here, a pattern of concentrations can be chosen, and then searched for in the dataset. Different correlation coefficients weight attributes of the data differently, e.g. magnitude of change matters for the Pearson

R, and direction of change for the Spearman rank correlation (Xia & Wishart 2016). To illustrate, high-low-high-low can be searched for with e.g. 3-1-2-1 or 2-1-2-1, where the first pattern emphasises the value of the first compared group.

4 Materials and methods

4.1 A cohort of mice

The FGF21 KO mice used in this study were created by PhD student Saara Forsström using the Cre-LoxP system. Deletor mice were crossed with previously characterised PGK-Cre (Lallemand et al. 1998), and *Fgf21*$^{LoxP/LoxP}$ (Potthoff et al. 2009) mice, resulting in the total body knock-out of FGF21 and yielding littermates of all the four different studied genotypes presented in Table 1.

The mice were euthanised at the age of 22-24 months and their tissues were collected for subsequent analysis. Snap frozen tissue samples of mouse *quadriceps femoris* (QF, skeletal muscle), *gastrocnemius* (GC, skeletal muscle), heart, liver, cerebral cortex, and dorsal and ventral hippocampi were used for the experiments presented in this thesis.

Strain name	Genotype
Deletor FGF21 KO	Twinkle$^{dup+/-}$; PGK-Cre; FGF21$^{flox/flox}$
FGF21 KO	PGK-Cre; FGF21$^{flox/flox}$
Deletor	Twinkle$^{dup+/-}$
Wild type	Twinkle$^{dup-/-}$

Table 1: The genotypes of the mice in the study.

4.2 Metabolomics

The metabolomic data set was measured by Vidya Velagapudi from QF samples with the same method as published in Nikkanen et al. (2016).

4.2.1 Selection of analytical tools and questions to guide the analysis

In bioinformatics it is possible to generate an endless amount of results in values, tables, matrices, lists of metabolites, all while losing sight of the underlying biology. While it is important to explore the data in various ways, an attempt at restraining the output of analyses was made in order to stay aware of what is relevant. Simple illustrations were drawn in R to satisfy a curiosity about individual points. Otherwise, a subset of analytical tools in MetaboAnalyst was picked for use, as presented in Table 4.2.1.

Analysis was done starting from the general and searching for details that could be explored further. To get started with data exploration, some initial questions were stated. As a reference, dot plots of all measured metabolites were drawn with t-test results printed alongside the plots. These plots and the tables of significances can be found in the Appendix.

Initial questions were:

> Which chemical compound classes are represented? That is, which metabolites does the instrumental setup allow to be measured? Does this introduce bias into which metabolites are represented?
>
> Which metabolic pathways are represented? How do the measured compounds scatter into metabolic pathways? By how many and which metabolites is each pathway represented?
>
> Are there any clearly visible trends on a big heatmap of all the data? Is there enough variance for separation of the groups with multivariate PC and PLS-D analysis?
>
> Are the same metabolites upregulated in the Deletor as previously published? Do the altered pathways behave in the same manner, also the enzymes involved?

Further metabolomic analysis expanded from this starting point. How do the metabolites relevant for the Deletor behave in both the KO mice? Where does the presence or absence of FGF21 make a difference? Do the FGF21 KO mice in WT background harbour alterations which are not observed in Deletor FGF21 KO mice?

4.2.2 Statistical preprocessing

The metabolomic data was provided in two different forms, first as original concentration values and second as values normalised against a reference sample as in Dieterle et al. (2006), and log-normalised. In the analysis using R, the original measured concentration values were used. For comparability with Nikkanen et al. (2016), outlier values were identified and removed using the robust regression and outlier removal (ROUT) method (Q = 1%) on GraphPad Prism 6.0 as in the article. In the analysis using MetaboAnalyst, the normalised data was used and no single outlier values were removed.

Two different tools were used for the metabolomic analysis. Fold change calculations and the assessment of statistical significance using Student's t-test were performed in R in order to generate

Method	Type	Format of the results
Heatmap with hierarchical clustering (HC)	Univariate and HC	Metabolites and/or samples organised by hierarchical clustering and concentration values plotted in a matrix on a colour scale
ANOVA with a post-hoc test: Fisher's least significant difference (LSD)	Univariate	p-values indicating the significance of difference between means of all genotype groups, and group pairs between which the significant difference is observed
Significance analysis of microarrays (SAM)	Univariate	p-values indicating statistically significant changes within this dataset
Random forests (RF)	Machine learning	Strength assessment of the classification tree forest, metabolites important for classification, and possible outlier samples
Principal component analysis (PCA)	Multivariate	Scores and loadings plots for selected dimensions of the created model
Partial least squares discriminant analysis (PLS-DA)	Multivariate	Scores and loadings plots for selected dimensions of the created model, and variable importance in projection scores for the metabolites
Correlation matrix with HC	Univariate and HC	Metabolites and/or samples organised by HC and correlation coefficients between metabolites plotted in a matrix on a colour scale
PatternHunter	Univariate	Plot of metabolites with the highest correlation with the queried pattern

Table 2: Some of the analytical tools available in MetaboAnalyst which were used in this study.

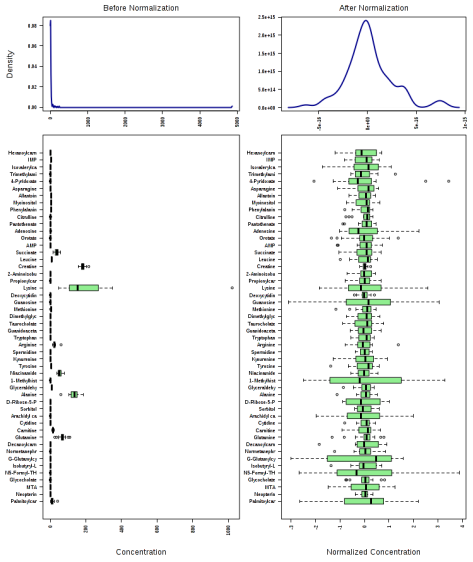


Figure 1: Illustrated data transformation and scaling of original concentration values in MetaboAnalyst. log transformation and mean centering of the Deletor and wild type genotypes.

barplots, dot plots, and volcano plots of the data. In the barplots, the data was chosen to be presented as fold changes with the standard error. In the dot plots, the original concentration values, the mean, and standard deviation are shown. All of the subsequent analysis was conducted in MetaboAnalyst. In addition to R and MetaboAnalyst, GraphPad Prism 6.0 was used only for the one method of outlier removal.

4.2.3 Data upload and missing value treatment

In R, the data was imported as comma separated values (csv) tables and handled as data frames. Means and fold changes were calculated, and the function "t.test" used for unpaired Student's t-tests (confidence level = 0.95) to compare the metabolite concentration means between all genotype pairs (Del/WT, Del/DelKO, KO/WT, DelKO/WT, DelKO/KO, and Del/KO).

The normalised concentration values were uploaded into MetaboAnalyst. Four missing values were left in this dataset, all in the sample AW3400D- (in 5,10-methenyl-THF, 5-methyl-THF, THF, and N5-formyl-THF). These were imputed by replacement with the column mean to allow for multivariate analysis. Replacement by a small value was also tried out. There, the assumption is that the missing concentrations were below the instrumental detection limit. This resulted in the sample with the missing values repeatedly showing up as an outlier in the analysis. The column mean worked better, and subsequent analysis looked reasonable.

For comparison between only a pair of genotypes at a time, the unnormalised original concentration values for each pair were uploaded into Metaboanalyst. Metabolites with ¿50% missing values were left out of analysis. Thereupon missing values were replaced by the column mean, and the data was log-transformed and mean centered. The result of this treatment for the genotype pair Deletor-WT is illustrated in Figure 1.

4.3 Detecting messenger RNA expression

qRT-PCR reactions were performed with samples from several mouse tissues to detect expression levels of mRNA for *Fgf21* and relevant enzymes. For this, total RNA was isolated from mouse liver, GC, and heart tissue samples. Additionally, RNA samples from mouse QF, cerebral cortex, and dorsal and ventral hippocampi isolated by the laboratory technician Tuula Manninen were used for analysis.

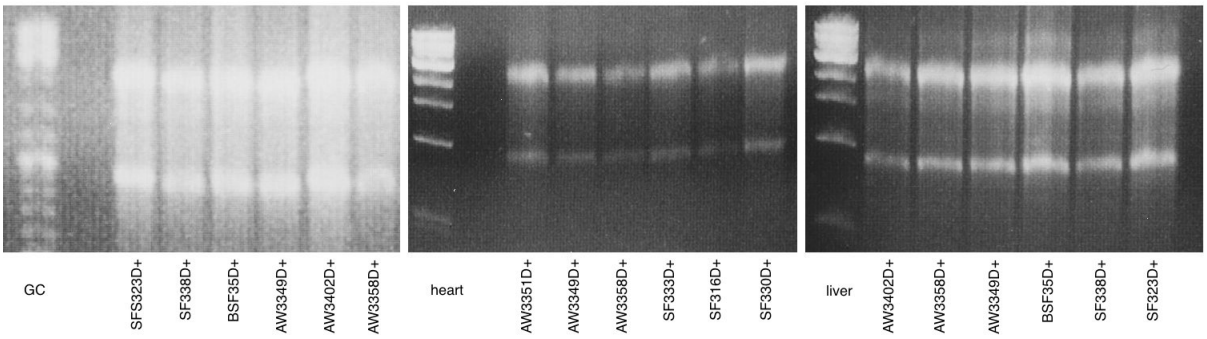


Figure 2: Total RNA isolated from *gastrocnemius*, heart and liver samples, with visible 18S and 28S ribosomal RNA bands.

4.3.1 Total RNA extraction

For GC, heart, and liver RNA, samples from Deletor and Deletor FGF21 KO mice were used. The samples were homogenised in the Trizol reagent (Invitrogen # 15596026) using a Precellys tabletop bead homogeniser for 30 s with the speed 5000. The total RNA was extracted in chloroform, precipitated with ethanol, and resuspended in TE buffer. RNA concentrations were measured with a NanoDrop1000 spectrophotometer.

The integrity of the isolated RNA was checked by agarose gel electrophoresis. A 1 w-% agarose gel with ethidium bromide was prepared. The samples were run on the gel at 90 V and 100 mA. 18S and 28S ribosomal RNA bands were observed under UV light in each sample on the gels, indicating that the isolated RNA was intact. The gels for GC, heart, and liver RNA can be seen in Figure 2.

4.3.2 cDNA synthesis

cDNA was synthesised from 500 ng of total RNA from each sample, using the Maxima first strand cDNA synthesis kit (ThermoScientific # K1671). Before synthesis the samples were treated with dsDNAse from the kit to minimise DNA contamination.

After cDNA synthesis, samples were selected to spot-check once more for RNA integrity by performing a PCR. Three samples were picked, one from each extracted tissue but otherwise randomly. The gene for mitochondrial helicase TWINKLE was selected for amplification, because for it primers were available which were complementary to locations in exons and thus suitable for the cDNA. The primers had been designed by PhD student Olesia Ignatenko.

cDNA corresponding to $\sim$13 ng total RNA was used for each reaction, along with primers presented

Gene	Forward primer	Reverse primer
Phgdh	gaccccatcatctctcctga	gcacacctttcttgcactga
Psat1	agtggagcgccagaatagaa	cttcggttgtgacagcgtta
Fgf21	ctgggggtctaccaagcata	cacccaggatttgaatgacc
β-actin	atgctccccgggctgtat	cataggagtccttctgacccattc
Twinkle	cgttttgaggacctgcctct	ttggacacctgcagataccg

Table 3: All primers used for polymerase chain reactions (PCR). *Phgdh* = phosphoglycerate dehydrogenase, *Psat1* = phophoserine aminotransferase, *Fgf21* = fibroblast growth factor 21.

in table 4.3.3, the Phire polymerase and 10 mM dNTP:s. The PCR products were separated by agarose gel electrophoresis on a 1.5 w-% gel. Bands of approximately the right size of 350 bp were observed, indicating that the RNA extraction had been successful, as can be seen in Figure 3.

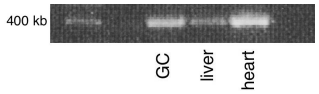


Figure 3: Products of the *Twinkle* polymerase chain reaction (PCR) from *gastrocnemius*, heart, and liver samples.

4.3.3 Quantitative PCR

qRT-PCR was performed on a BioRad CFX96 Touch Real-Time PCR System using the SYBR Green detection system. iQ SYBR Green Master mix (BioRad # 4312704), forward and reverse primers, and cDNA sample were added into 20 μl reactions in each well. cDNA corresponding to ∼8 ng total RNA was used per well for all runs and samples were run in triplicates, except for *Fgf21*. For detecting *Fgf21*, cDNA corresponding to ∼80 ng total RNA was used and samples were run in duplicates. *β-actin* was used as the housekeeping gene control. The results were calculated using the comparative C_T method (Schmittgen & Livak 2008). All used primers are presented in Table 4.3.3.

4.4 Detecting protein expression

The expression levels of selected proteins were measured in QF samples using western blotting.

Type	Host	Anti	Supplier	Cat. No.	Dilution
primary	rabbit	CTH	Proteintech	12217-I-AP	1:2000
primary	rabbit	GCLC	Proteintech	12601-1-AP	1:1000
primary	chicken	MTHFD2	Abcam	ab37840	1:2000
primary	rabbit	MTHFD1L	Proteintech	16113-1-AP	1:1000
secondary	goat	mouse	Jackson ImmunoResearch	115-035-146	1:10 000
secondary	goat	rabbit	Jackson ImmunoResearch	111-035-144	1:10 000

Table 4: Used antibodies and their dilutions. All dilutions were made with 5 % milk in tris-buffered saline with Tween 20 (TBST), except for anti-CTH, which was diluted in 3 % bovine serum albumin (BSA) in TBST. CTH = cystathionine gamma-lyase, GCLC = glutamate-cysteine ligase, MTHFD2 = methylene tetrahydrofolate dehydrogenase, MTHFD1L = C1-tetrahydrofolate synthase.

4.4.1 Preparation of total protein lysates

The total protein was extracted from QF samples of mice of all four studied genotypes. 20–30 mg samples from each mouse were homogenised in TBS using the Precellys bead homogeniser at speed 2500 rpm. The samples were placed on ice for 30 s between 30 s rounds of homogenisation and these steps were repeated until no more intact muscle tissue was visible in the tube (5–6 times). 1 % Triton-X was added to each tube. The samples were left on ice for 30 min and then spinned at full speed at +4°C. The supernatants were kept while discarding the pellets.

The protein concentrations of the lysates were determined using the Bradford assay (Bradford reagent BioRad # 5000006). BSA standards diluted from a stock solution were used as reference. The standards were reutilised from a BCA assay kit (Thermo Fisher # 23225). Sample lysates were diluted 1:20 and 10 μl of diluted sample was added to 180 μl Bradford reagent per well on a 96 well plate and incubated for 5 minutes. The plate was read at 595 nm on a spectrophotometer and total protein concentrations were determined based the resulting calibration curve.

4.4.2 Gel electrophoresis and immunoblotting

The total protein samples were buffered with Laemmli. 15 μg total protein was loaded into each well of a 4–20 % gradient SDS-polyacrylamide gel (BioRad # 4568095). 10 μl of the Dual Color protein standard (BioRad # 1610394) was loaded into one well for reference. The gel was run at 60 mV until the bands of the standard separated and at 120 mV until completion. The separated proteins were transferred onto a PVDF membrane using preassembled packs (BioRad # 1704156) on BioRad's Turbo Transfer System. The membrane was imaged BioRad's ChemiDoc XRS+ apparatus to detect total protein.

The membrane was blocked in 5 % milk in TBST for 1 h, followed by an overnight incubation with

a primary antibody, and a 1 h incubation with a secondary antibody conjugated with horseradish peroxidase (HRP). The used antibodies and their respective dilutions and solvents are listed in Table 4. The membrane was kept under agitation at all times and washed with the antibody solvent between steps for 3*5 min. An enhanced chemiluminescence (ECL) reagent (Clarity™ Western ECL Substrate BioRad # 1705061) was added and the membrane was imaged again with the ChemiDoc XRS+ to detect antibody-specific protein bands.

The western blots were quantified using ImageJ as in Miller (2010) and normalised against total protein. For the normalisation, a uniform area not containing the protein of interest was selected in the imaged stain-free gel showing total protein. The areas used for normalisation of each blot are shown in Figure 13 alongside the Western Blot results. Though the results were quantified, statistics were not calculated because only three samples of each genotype were measured.

5 Results

5.1 Verification of the knock-out

Expression levels of *Fgf21* mRNA were measured in a panel of tissues from Deletor and Deletor FGF21 KO mice in order to ascertain that the global knock-out of FGF21 had been successful. cDNA from samples of GC, heart, liver, cerebral cortex, and dorsal and ventral hippocampi were used.

In all of the brain samples, *Fgf21* was not detected. GC, heart, and liver showed very low expression in the Deletors but not in the Deletor FGF21 KO tissues. A Deletor QF positive control showed clear expression 10 cycles earlier than the Deletor GC, heart, and liver samples.

Because most of the experiments had a negative result and any expression was only detected at very late PCR cycles, the results are not presented quantified and as a graph. The essential results are: first, that the knock-out of FGF21 had been successful and second, that low expression of *Fgf21* in the Deletor GC muscle exists.

5.2 A rough large scale pattern in metabolite levels

A heatmap of all concentration values in the data set, as measured by our collaborator Vidya Velagapudi, was generated and can be seen in Figure 4. The metabolites were organised using hierarchical clustering with the Pearson R coefficient as the distance measure and Ward as the clustering algorithm. Care was taken not to cluster the samples and instead keep the groups intact, so that possible trends specific for genotype would become apparent.

A rough general pattern to metabolite concentration values can be seen in the heatmap, when the lower large branch of the hierarchical clustering tree is scrutinised. In this subset of metabolites, Deletors have high concentration values and FGF21 KO mice have low ones. Deletor FGF21 KO mice settle in between the former two and WT mice have variable values.

No obvious outlier samples are observed. Many of the low data values are found to be concentrated into single samples (e.g. BSF34D- and AW3360D+), even though the data has been thoroughly normalised. One sample has a set of noticeably high metabolite concentrations (SF324D-). None of these samples is aberrant enough for confident classification as an outlier.

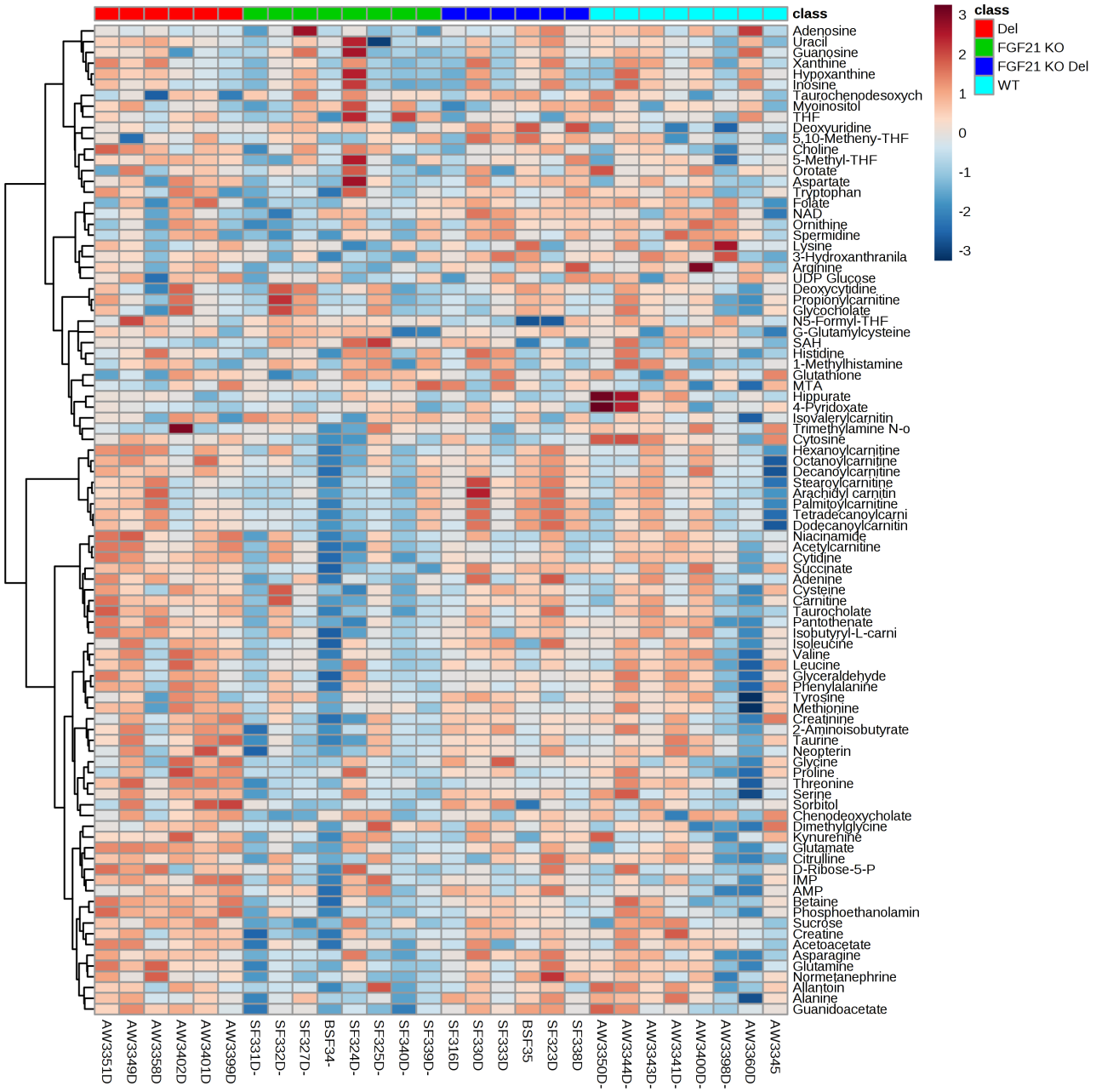


Figure 4: A heatmap of the concentration values of all analysed metabolites.

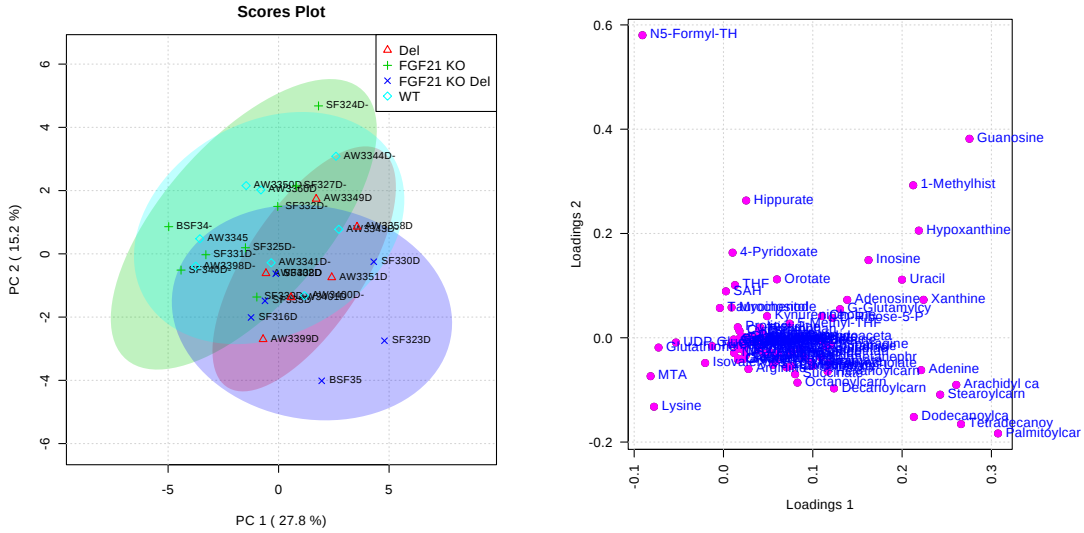


Figure 5: Principal component analysis (PCA). (Left) The scores plot of the first two principal components and (right) the corresponding loadings plot.

5.3 No separation of the genotypes using multivariate methods

5.3.1 Analysis of all genotypes together

The PCA scores plot of the first two principal components does not show definitive separation between any of the groups, as can be seen in Figure 5. An imprecise separation can be observed on the up-left to down-right diagonal axis in the plot. Based on the observation of this axis and the loadings plot respective to the scores plot, a group of carnitines could be important for the Deletor FGF21 KO phenotype (the arachidyl, stearoyl, dodecanoyl, tetradecanoyl and palmitoyl carnitines). Accordingly, N5-formyl-THF, hippurate, and 4-pyridoxate are probably important for the WT and FGF21 KO phenotypes.

Similar to the PCA, there is no clear separation of the groups in PLS-DA (Figure 6). Again, hippurate and 4-pyridoxate might be important for the FGF21 KO and WT phenotypes. In the respective loadings plot, a cluster of carnitines appears again, this time together with xanthine and 1-methylhistamine, all of which have thus similar loading scores. The fifteen most important metabolites in forming this PLS-DA model are listed in Figure A2. The three most important ones, hippurate, 4-pyridoxate, and N5-formyl-THF also attracted attention in the loadings plots for both PLS-DA and PCA.

Cross validation and permutation test scores for the PLS-D model can be seen in the Appendix in

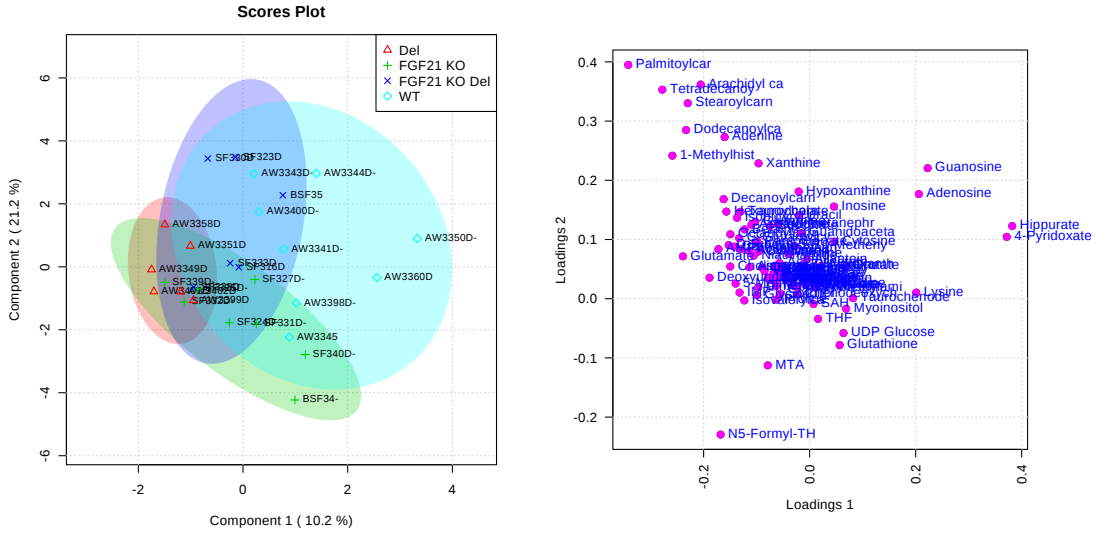


Figure 6: Partial least squares discriminant analysis (PLS-DA). (Left) The scores plot of the first two components and (right) the corresponding loadings plot.

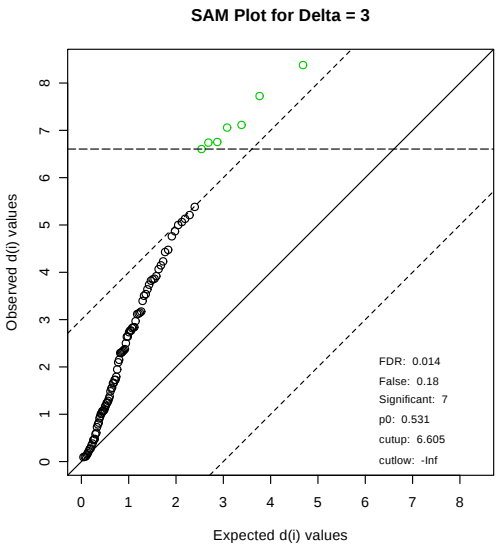
Figure A1. Based on the cross validation, the two component model is most optimal for this data set, even though its Q-value is negative, which indicates a bad fit. The permutation test gives a p-value of 0.13 for the prediction accuracy of the model (2000 permutations). The cross validation and permutation test together indicate, that the PLS-D model is not strong. Based on this analysis, there is no clear large scale or global difference between all the four genotype groups when they are analysed together.

5.3.2 Pairwise comparison of genotypes

PCA could not classify genotype groups when only two genotypes were analysed at a time. PLS-DA could, but the differences indicated by loadings plots were in metabolites that did not give any coherent further information, as discussed later in more detail.

5.4 Specific metabolic changes differ between genotypes

The three analytical methods SAM (significance analysis of microarrays), ANOVA, and RF together pintpoint eight metabolites, which provide separation between the genotypes in this study. All seven of these metabolites are closely related to the one carbon cycle metabolic pathway, implying that FGF21 modifies 1C metabolism. Below, the analysis leading to these specific metabolites is presented.



Number of trees to grow: 2000 Submit

Number of predictors to try for each node: 7

The OOB error is 0.5

	Del	FGF21 KO	FGF21 KO Del	WT	class.error
Del	3	0	1.0	2.0	0.5
FGF21 KO	0	6	1.0	1.0	0.25
FGF21 KO Del	1	0	3.0	2.0	0.5
WT	3	3	0.0	2.0	0.75

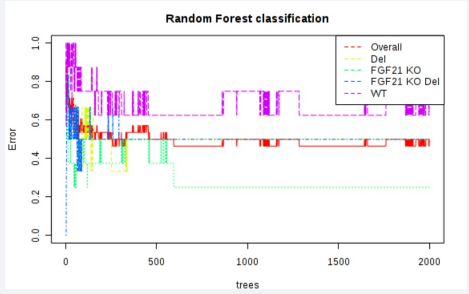


Figure 7: (Left) Significance analysis of microarrays (SAM) with the false discovery rate (FDR) controlled by a strict delta = 3. The analysis gives 7 significant metabolites, out of which 0.18 could be false with the FDR 0.014. (Right) RF error rate calculations. Overall and class error rates are plotted over the number of generated classification trees.

5.4.1 Filtering out eight interesting metabolites

In the SAM analysis, different delta values for controlling the false discovery rate (FDR) were tried out. The strictest obtained result was with delta = 3, where the seven metabolites succinate, glutamate, betaine, phosphoethanolamine, deoxyuridine, creatine and guanidinoacetate were found to be significantly changed with a FDR of 0.18, as can be seen in Figure 7.

One-way ANOVA with Fisher's LSD post-hoc tests was performed. 32 metabolites had a p-value <0.05, meaning that in these metabolite hits there is at least one genotype group pair which differs with that level of significance. The post-hoc test reveals the group pairs in which the difference is observed. The seven highest ranking metabolites completely coincide with the results from SAM, as presented in the summary in Table 5.4.1. The 15 most significant metabolites from the ANOVA analysis and their respective pairs with significant changes are listed in the Appendix in Table A1.

The random forest (RF) algorithm was run with the parameters "2000 trees" and "7 predictors per node". Important metabolites based on this analysis are listed in Table 5.4.1 alongside ANOVA and SAM values, and also in Figure A2 next to metabolites important in the PLS-D model. As can be seen in both figure and table, the important metabolites based on PLS-DA and RF do for the most part not agree. Deoxyuridine, glutamate, and cytosine come up in both analyses.

	ANOVA	p-value	SAM	raw p	RF	MDA	PLS-DA	VIP
1	Succinate	0.0005	Succinate	0.0005	Succinate	0.011	Hippurate	3.7
2	Glutamate	0.0009	Glutamate	0.0009	Creatine	0.009	4-Pyridoxate	3.2
3	Betaine	0.0014	Betaine	0.0013	Deoxyuridine	0.009	N5-Formyl-THF	2.4
4	PE	0.0015	PE	0.0013	Threonine	0.009	Adenosine	2.3
5	Deoxyuridine	0.0018	Deoxyuridine	0.0016	Betaine	0.008	Lysine	2.2
6	Creatine	0.0019	Creatine	0.0017	Glutamate	0.008	Guanosine	2.2
7	GAA	0.0021	GAA	0.0020	PE	0.006	Choline	2.2

Table 5: Comparison of metabolite hits. ANOVA = analysis of variance, SAM = significance analysis of microarrays, RF = random forests, PLS-DA = partial least squares discriminant analysis, MDA = mean decrease accuracy, VIP = variable importance in projection, PE = phosphoethanolamine, GAA = guanidinoacetate, THF = tetrahydrofolate.

The error values and the corresponding stabilisation plot for the algorithm can be seen in Figure 7. An OOB overall error rate of 0.5 was reported, which is quite high. The FGF21 KO group had the lowest class error rate of 0.25, which became stable already at around 600 trees, whereas the WT group had a very high class error rate of 0.75, which did not fully stabilise even after 2000 trees. The Deletor and Deletor FGF21 KO groups had error rates of 0.5, conforming with the overall OOB error rate. Based on these error values, there is most variation between samples in the WT group and least in the FGF21 KO group.

5.4.2 Observed patterns reflect known upregulation in the Deletor

The top seven metabolites from SAM and ANOVA coincide completely, and RF nearly equally, adding only threonine to the list as seen in Table 5.4.1. These eight metabolites are shown in Figure 8 in dot plots. To complement the three statistical methods already used, significances between genotypes are shown as determined by Student's t-test. All of these metabolites are closely related to 1C metabolism.

Glutamate, betaine, phosphoethanolamine, and threonine act in a similar way: they have elevated levels in the Deletor, which are rescued by the knock-out of FGF21. Choline and acetylcarnitine share their concentration pattern with these three, and they are shown in Figure 5.4.2. Deoxyuridine acts unlike all other metabolites in that it has elevated levels in all genotypes but the wild type. Removal of FGF21 leads to elevated levels of deoxyuridine.

Finally, succinate, guanidoacetate, and creatine have low levels in the FGF21 knock-out mice without the Deletor mutation. This implicates, that the Deletor pathomechanism keeps these metabolites from decreasing in the Deletor FGF21-KO mice. Metabolic changes in the Deletor in general

include high levels of metabolites, so presumably high levels of other metabolites in the intertwined metabolic network keep these three from lowering in the Deletor FGF21-KO.

5.5 Fibroblast growth factor 21 adds to the disease progression

As already seen in the individual metabolites glutamate, betaine, phosphoethanolamine, and threonine, aspects of the pathology in the Deletor tend to be rescued by the knock-out of FGF21. This is discovered further in another look at the metabolomic data, and in expression levels of specific enzymes as examined by mRNA and protein measurements.

5.5.1 Upregulated metabolites tend to be rescued

To illustrate significantly changed metabolites, bar plots, volcano plots, and dot plots were drawn in R. These illustrations were generated from the unnormalised but outlier-removed data, and significances were determined using Student's t-test.

Of these, a plot depicting upregulated metabolites in the Deletor revealed their consistent dependence on FGF21 (Figure 10). Here, all significantly changed metabolites of the Deletor compared to WT were plotted as fold changes, and the same metabolites picked from the Deletor FGF21 KO data for comparison, again fold changes compared to WT. Strikingly, most of these metabolites returned to WT levels in the Deletor FGF21 KO mice, implying that removal of FGF21 rescues part of the phenotype in the Deletor.

In the volcano plots in Figure 11 the plot of the Deletor shows mostly upregulated metabolites and the FGF21 KO mostly downregulated metabolites, which is the same rough large scale pattern as observed in the heatmap in Figure 4. The significantly changed metabolites in the Deletor FGF21 KO are mostly not the same as in the Deletor, which is natural, given the rescue observed above. This does however point out, that the removal of FGF21 causes also other effects than solely rescuing the upregulation of metabolites in the Deletor.

5.5.2 Corresponding changes on enzyme and metabolite levels

The serine pathway channels glucose carbon units into 1C metabolism and it has been shown to be upregulated in mitochondrial myopathy in the Deletor and in human patients (Tyynismaa et al. 2010, Nikkanen et al. 2016). Elevated glucose uptake into the diseased Deletor muscle feeds serine into the folate cycle, where one carbon units are converted into formate and used for methylation

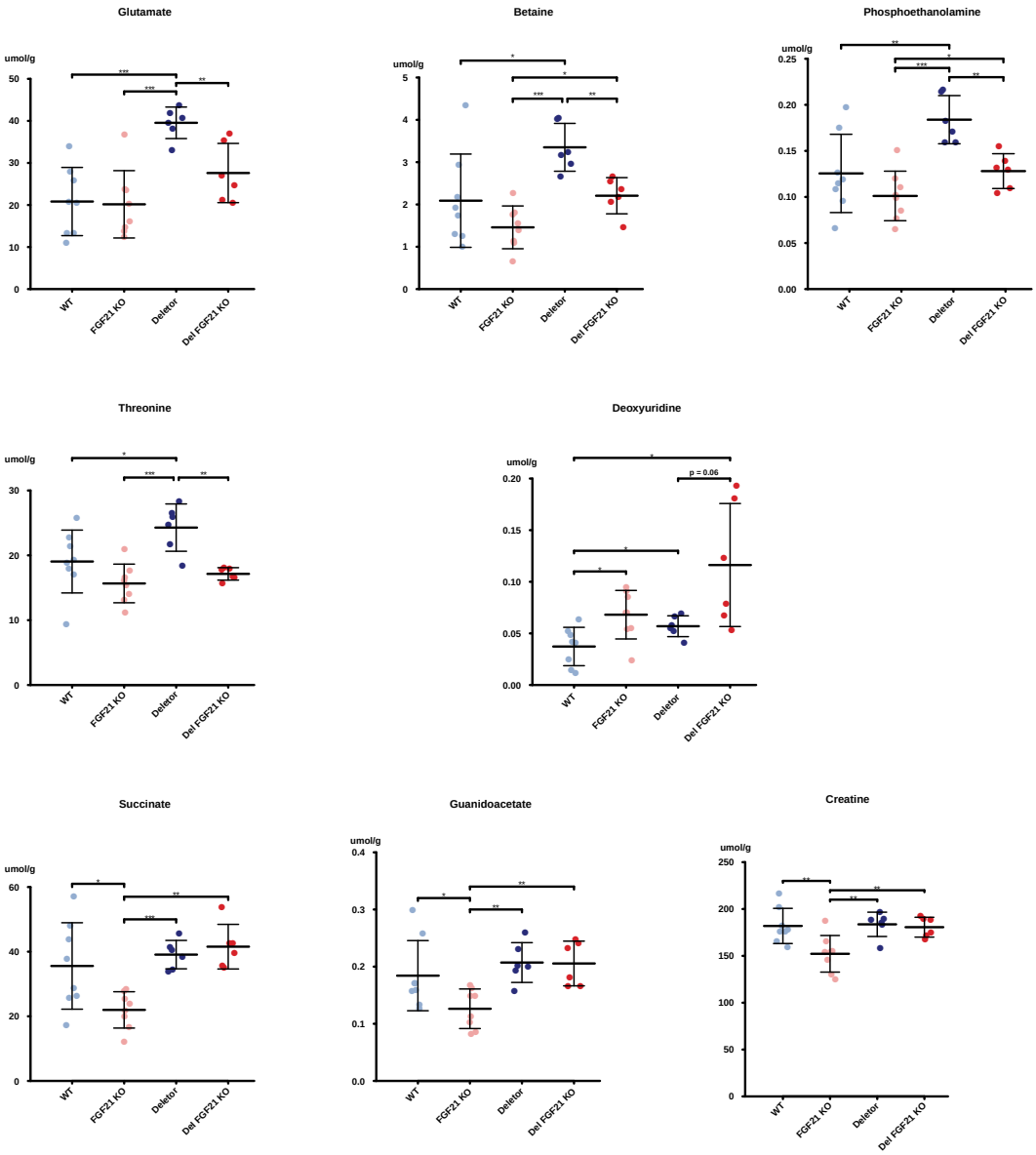


Figure 8: Dot plots of selected metabolites, which frequently turn up in the analysis. The significances between each pair of groups were determined by Student's t-test. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

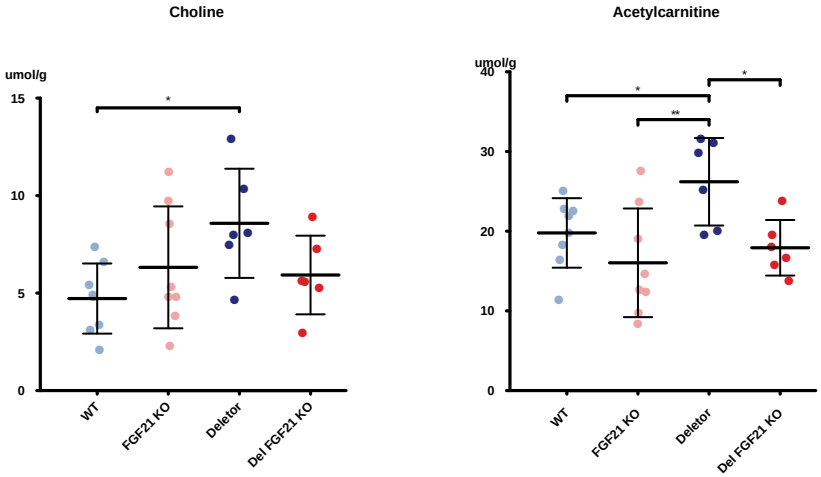


Figure 9: Choline and acetylcarnitine levels shown as dot plots. The significances were determined by Student's t-test, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

reactions, and into the transsulfuration pathway that produces glutathione. Out of these pathways, enzymes shown to be upregulated in the Deletor were selected for study to see if removal of FGF21 has an effect on them. As on the metabolite level, FGF21 drives the upregulation of part of the studied metabolic pathways on the enzyme level.

The mRNA expression of two enzymes from the serine de novo synthesis pathway were measured in all four genotypes. These were D-3-phophoglycerate dehydrogenase (*Phgdh*) and phosphoserine aminotransferase (*Psat1*), both of which are upregulated in the Deletor (Tyynismaa et al. 2010, Nikkanen et al. 2016). The results are shown in Figure 12. The expression of both enzymes was high in the Deletor and both were brought down in the absence of FGF21. In *Phgdh* the effect was evident in both WT and Deletor backgrounds.

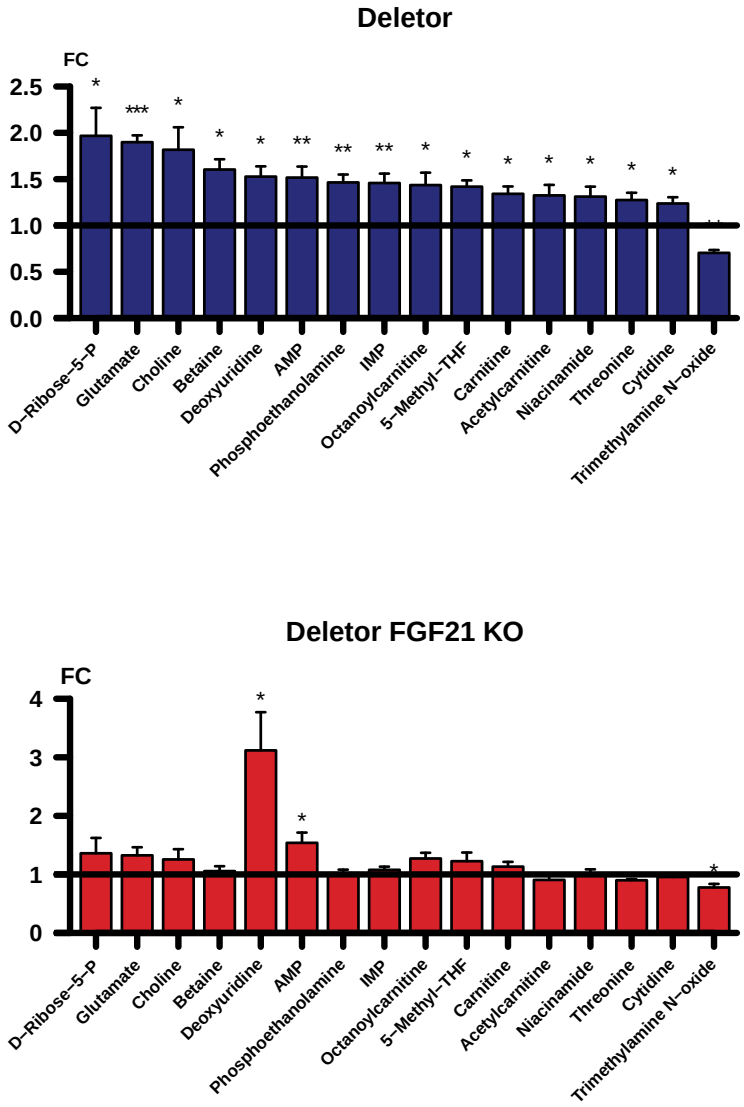


Figure 10: All metabolites that are changed significantly in the Deletor compared to WT, and the same metabolites in the Deletor FGF21 KO. Fold changes of both genotypes were calculated compared to WT. The significances between groups were determined by Student's t-test. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

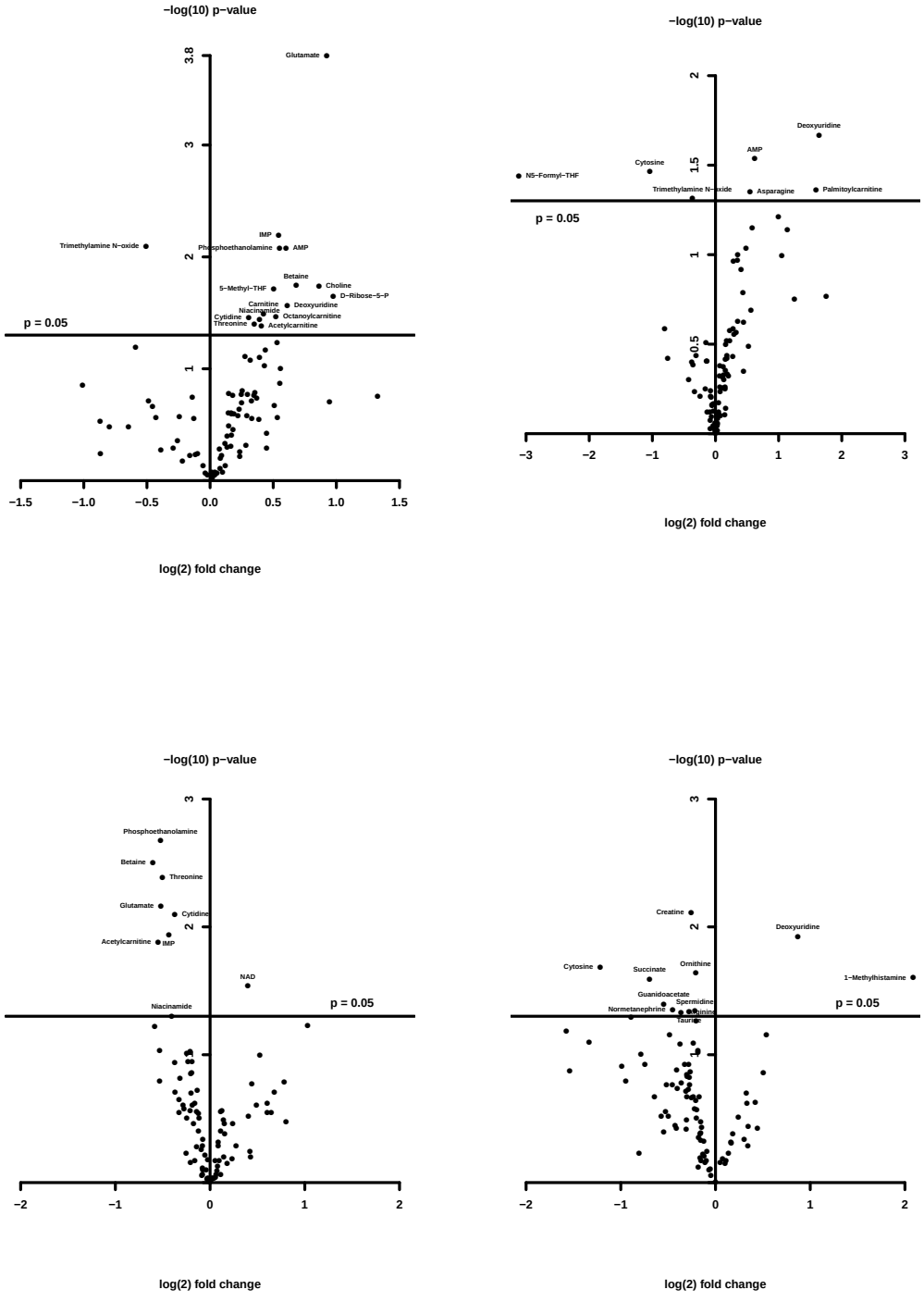


Figure 11: Volcano plots. (Up-left) Deletor vs WT. (Up-right) Deletor FGF21 KO vs WT. (Down-left) Deletor FGF21 KO vs Del. (Down-right) FGF21 KO vs WT. The rough pattern seen in the global heatmap repeats itself in the volcano plots: the Deletor has mostly high levels of metabolites and the FGF21 KO mostly low levels of metabolites. It can also be seen, that the significantly changed metabolites in the Deletor FGF21 KO are mostly not the same as in the Deletor.

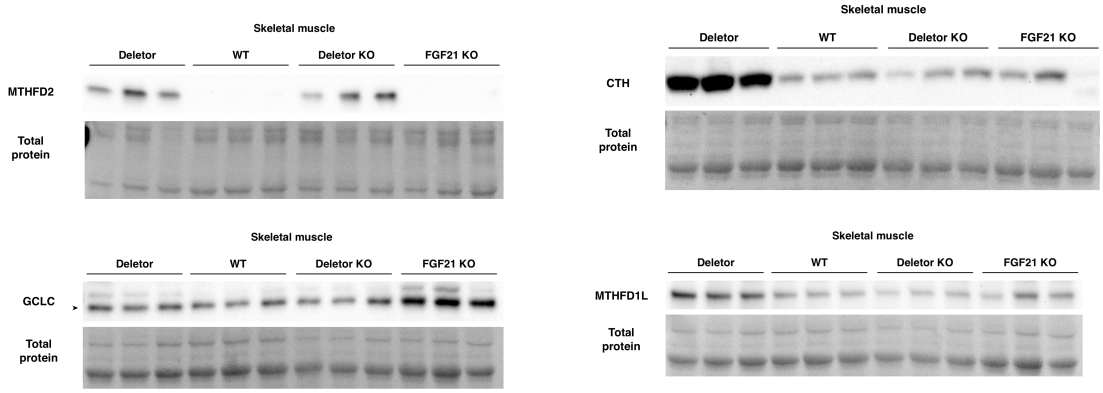


Figure 13: Western blots along with the total protein area used for quantitation. Expression in skeletal muscle. (Up-left) MTHFD2 = methylene tetrahydrofolate dehydrogenase. (Up-right) CTH = cystathionine gamma-lyase. (Down-left) GCLC = glutamate-cysteine ligase. (Down-right) MTHFD1L = C1-tetrahydrofolate synthase.

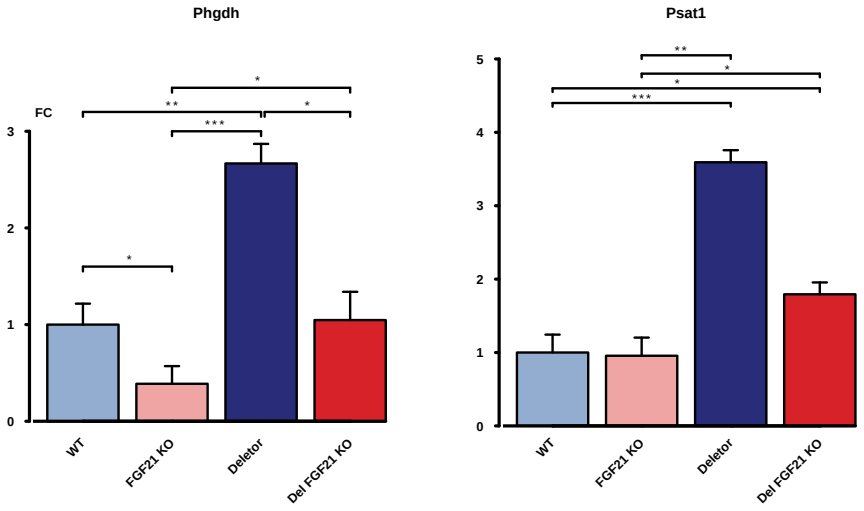


Figure 12: qRT-PCR in QF skeletal muscle. Data presented as fold change compared to WT. *Psat1* = phosphoserine aminotransferase, *Phgdh* = D-3-phosphoglycerate dehydrogenase. The significances between each pair of groups were determined by Student's t-test. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

The enzymes Methylene tetrahydrofolate dehydrogenase (MTHFD2), C1-tetrahydrofolate synthase (MTHFD1L), cystathionine gamma-lyase (CTH) and glutamate-cysteine ligase (GCLC) were studied by western blotting (Figure 13). MTHFD2 and MTHFD1L catalyse the interconversion of folate intermediates in the folate cycle. CTH and GCLC are enzymes of the transsulfuration pathway.

GCLC is upregulated in FGF21 KO mice compared to the other genotypes. The upregulation in the Deletor is not replicated in a clear manner. The enzyme catalyses the synthesis of gamma-glutamylcysteine out of glutamate and cysteine. The respective metabolites were not found to be altered, except for the upregulation of glutamate in the Deletor. The metabolite measurement for gamma-glutamylcysteine had very low precision. It is difficult to draw conclusions regarding the enzyme GCLC in this context.

CTH and MTHFD1L expression return to WT levels in Deletor FGF21 KO mice. MTHFD2 is slightly less highly expressed in Deletor FGF21 KO mice than in Deletors based on the quantitations of the blots (Figure 14), but essentially it is not regulated by FGF21 based on this result.

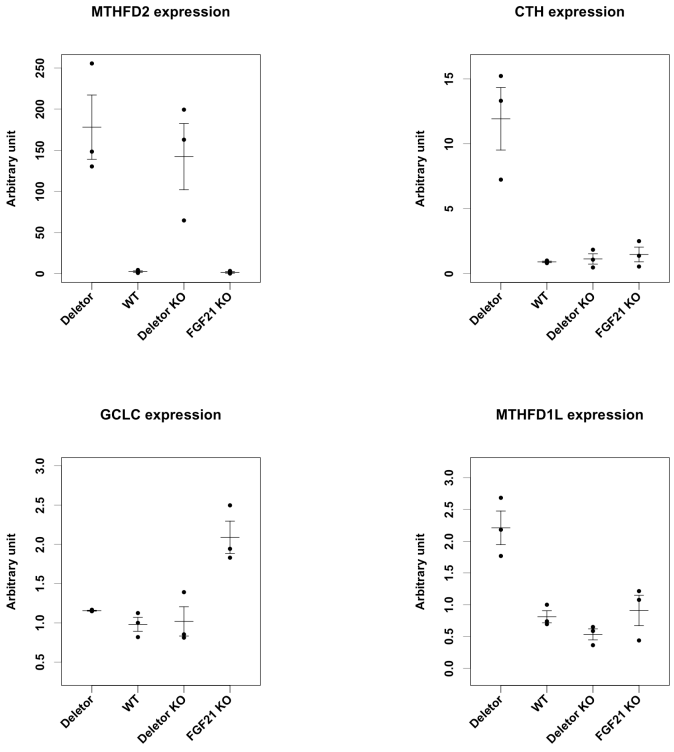


Figure 14: Western blot quantitations. (Up-left) MTHFD2 = methylene tetrahydrofolate dehydrogenase. (Up-right) CTH = cystathionine gamma-lyase. (Down-left) GCLC = glutamate-cysteine ligase. (Down-right) MTHFD1L = C1-tetrahydrofolate synthase. No statistics were calculated because the number of samples per group was only three.

5.6 Long-chain acylcarnitines are not affected

Acylcarnitines were analysed, because they have high levels in PEO patient muscle (Nikkanen et al. 2016). The results of this KO study indicate, that long-chain acylcarnitines are not regulated by FGF21, as can be seen in the heatmaps of the top 25 metabolites from the ANOVA and RF analyses are shown in Figure 15. These subsets of metabolites follow the trends observed above. Deletors tend to have increased concentration values, and FGF21 KO mice tend to have low concentration values. WT and Deletor FGF21 KO mice settle in between these extremes.

Their concentrations seem to act approximately in concert in all of the different genotypes, and the hierarchical clustering algorithm has accordingly placed these metabolites closeby each other. The acylcarnitines are high in both the Deletor and Deletor FGF21 KO mice, and low in the other two groups, thus unaffected by the KO of FGF21 so their upregulation is caused by something other than FGF21 in the disease mechanism. It is possible that these compounds stand out because their class is well represented in the data set. Other compounds in the heatmaps in Figure 15 represent various different metabolic pathways.

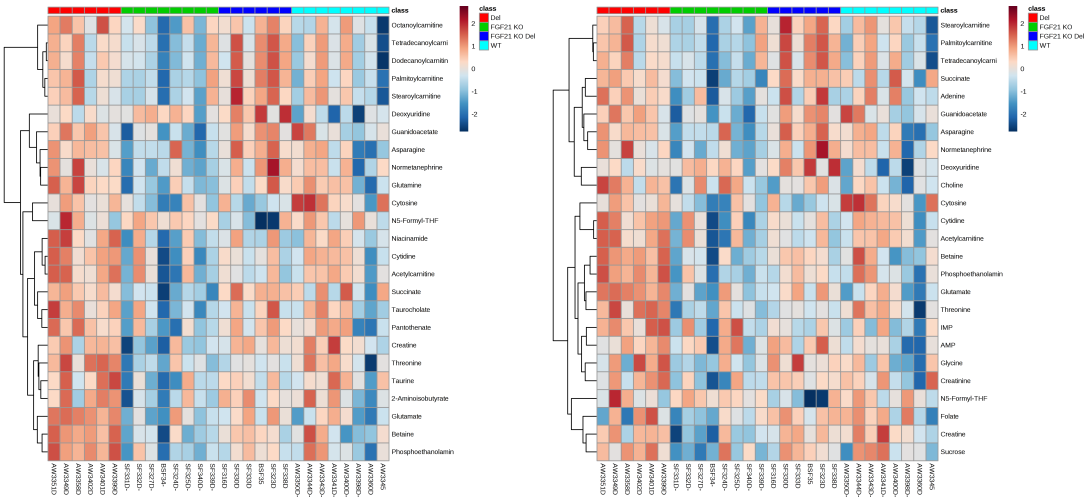


Figure 15: Heatmaps of the top 25 metabolites from (left) ANOVA and (right) RF analysis. The metabolites are organised using hierarchical clustering.

6 Discussion

The results of this study comprise that FGF21 drives some but not all of the alterations on mRNA, protein, and metabolite levels in mitochondrial myopathy as seen in the Deletor. FGF21 does this in a widespread manner, while still not being the only driving force in the disease. It also plays a role in normal physiology, as revealed by changes in the FGF21 KO metabolite levels. The alterations seen in this study revolve around closely related pathways to such extent that most altered metabolites and some enzymes could be illustrated in compact form in the cartoon in Figure 16.

FGF21 is expressed in a variety of tissues, has a large variety of actions and is induced via many different signals. Its numerous functions are reflected by the fact that its removal caused many alterations in the KO mice of this study. The identified alterations can be grouped under larger themes as follows. Alongside these points on physiology some technical aspects on metabolomics are discussed.

6.1 Physiology

6.1.1 Regulatory hierarchy

The physiology of FGF21 is complex, as is reviewed from a general viewpoint in Fischer & Marator-Flier (2016), and from a viewpoint of stress signalling in Salminen et al. (2017). To illustrate this statement, it was recently established that mTORC1 regulates FGF21 (Khan et al. 2017), and that FGF21 contributes to mTORC1 activation (Minard et al. 2016). It is also known, that ATF-group transcription factors regulate transcription of AARE genes, among them FGF21 (Tyynismaa et al. 2010), but the study in this thesis shows that FGF21 also regulates transcription of genes that seem to be under AARE influence (*Phgdh* and *Psat1*).

An essential finding in this study is a generalisation of that last point. In the signalling direction mTOR1 – ATF4 – FGF21, the hormone forms its own feed-back regulatory step in the tissue in an autocrine manner, influencing the expression of its peer genes under promoters with AARE sequences.

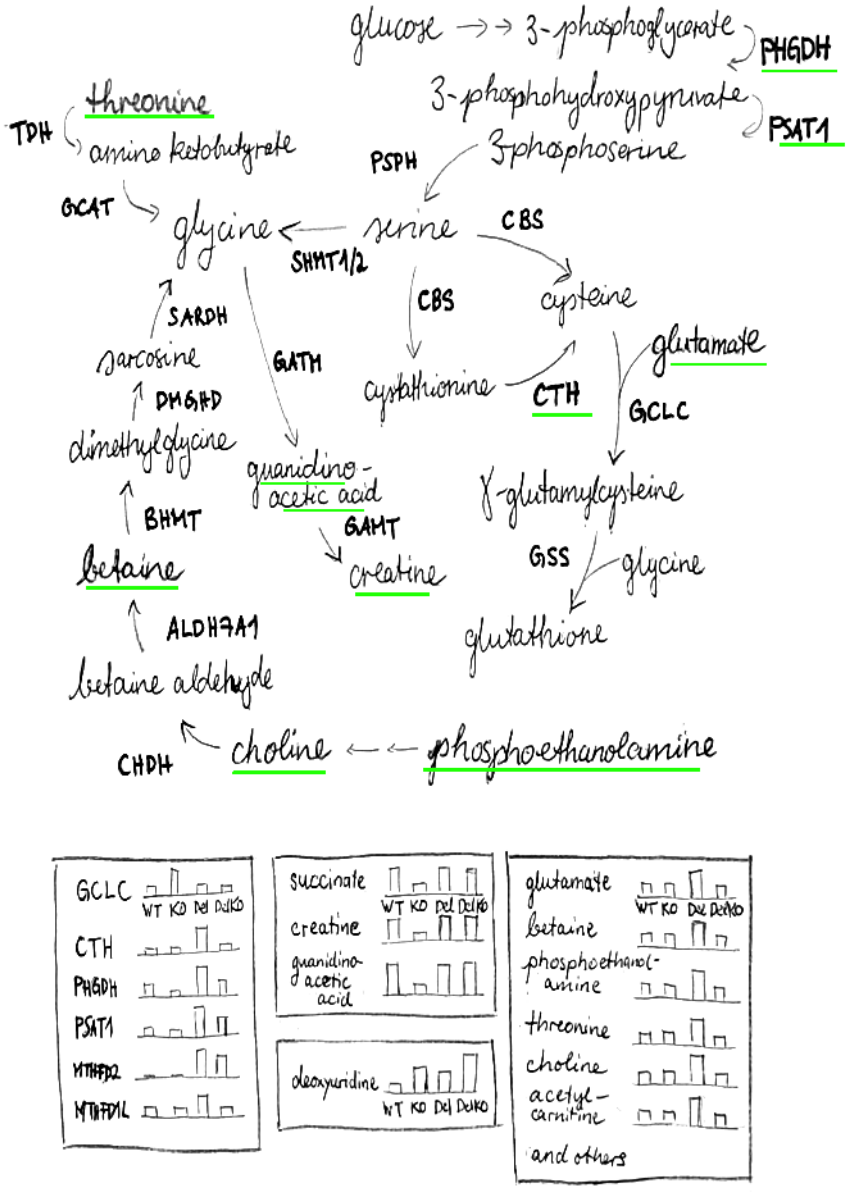


Figure 16: A cartoon that summarises the essential results from this study and shows the featured metabolites and enzymes in a metabolic pathway context. In green are metabolites and enzymes in which removal of FGF21 caused an effect in this study. Above: metabolic pathways from *Mus musculus* (Mouse). Not all results of this study are depicted here. Below: caricatured graphs of interesting results. See the Results-section for the respective accurate graphs. Genotypes are always in the order WT - FGF21 KO - Deletor - Deletor FGF21 KO. The enzymes listed in the graphs below on the left are GCLC = glutamate-cysteine ligase, CTH = cystathionine gamma-lyase, PHGDH = D-3-phosphoglycerate dehydrogenase, PSAT1 = phosphoserine aminotransferase, MTHFD2 = methylene tetrahydrofolate dehydrogenase, and MTHFD1L = C1-tetrahydrofolate synthase. Other enzymes mentioned in the pathways above catalyse the converting reactions between the respective depicted metabolites.

6.1.2 One carbon metabolism - carbon donors

The perhaps most essential finding of this study is that FGF21 drives the compensatory upregulation of alternative 1C donors in the face of the impaired formate production in mitochondrial myopathy2002. Low formate production concurs with mitochondrial dysfunction as reported in Bao et al. (2016), Nikkanen et al. (2016). In the diseased tissue there is an attempt to balance impaired formate production by upregulation of serine de novo synthesis genes via ATF4. The alternative carbon donors upregulated by FGF21 in the Deletor found in this study are choline, betaine, and threonine, which can be seen in the cartoon in Figure 16. Threonine feeds into 1C metabolism in mice, where it is converted by threonine 3-dehydrogenase (TDH) to amino ketobutyrate and by 2-amino-3-ketobutyrate coenzyme A ligase (GCAT) into glycine. TDH is inactive in humans (Edgar 2002).

The shuffling of 1C units is mediated by the folate cycle. The impact of the knock-out of FGF21 on folate cycle intermediates did not yield many tangible results. The measured folate cycle intermediates are illustrated in Figure A3 in the Appendix. The upregulation of these intermediates as reported in Nikkanen et al. (2016) was mostly not replicated. The only differences are in N5-formyl-THF, where the Deletor FGF21 KO has notably low levels of the metabolite. Most of the measurement points for this metabolite were spread out with low accuracy, and it is easiest to explain the result as an unsuccessful measurement.

Based on the western blot analysis, the folate cycle enzyme MTHFD2 that is strongly upregulated in the Deletor, supporting previously published results. Its expression did not change with the KO, so it probably is not controlled by FGF21 but instead independently of it. The enzyme is often seen upregulated in cancers (Nilsson et al. 2014), so there must be some strong mechanisms related to growth and proliferation in which it is involved, which are outside of the influence of FGF21.

Choline, serine, and glycine are quantitatively the most important dietary sources of 1C units for the folate cycle (Ducker & Rabinowitz 2017). Based on Student's t-test choline is also significantly high in the Deletor (Figure 5.4.2) and has the same rescue pattern as glutamate, betaine, PE, and threonine, so it was added to the cartoon in Figure 16 alongside the other essential metabolite hits. It would have been interesting to see whether FGF21 also influences the levels of serine and glycine, but the upregulation of these was not replicated in the metabolomic measurement of this study and differences between genotypes were not seen.

6.1.3 Deoxyuridine and amino acids

1C metabolism supports the biosynthesis of purines and thymidine. Deoxyuridine is a thymidine precursor. In this study it behaved unlike most metabolites in that it was upregulated in all three groups, even higher in the Deletor FGF21 KO than in the other two. The change in the Deletor FGF21 KO was significant in the post-hoc test of ANOVA, and had a p-value of 0.06 in the t-test. These results implicate that FGF21 always acts to keep deoxyuridine levels low, which is emphasised in the myopathy, because the disease state seems to drive the levels higher.

Amino acids have generally been reported high in the Deletor (Ahola-Erkkilä et al. 2010, Tyynismaa et al. 2010, Nikkanen et al. 2016). FGF21 seems to not play a role in this tendency. Glutamate and threonine levels did change with the knock-out, but to pick an example, asparagine dit not. Asparagine synthetase (ASNS) is one of the proteins upregulated in the Deletor by the same AARE as FGF21 (Tyynismaa et al. 2010). Here, the amino acid was significantly upregulated in the Deletor FGF21 KO with a FC of 1.5 (non-significant in the Deletor, FC 1.3), which conforms with the regulation of ASNS commencing independently of FGF21.

Glutamate has been shown to accumulate within the mitochondrial matrix in mitochondrial dysfunction of different causational origins (Chen et al. 2016). The fact that the FGF21 deletion in this study rescued glutamate levels in particular further points in the direction tha FGF21 is an essential part of the Deletor pathomechanism.

6.1.4 Creatine synthesis and energy metabolism

Creatine, guanidinoacetate, and succinate levels were significantly decreased in the FGF21 KO samples compared to all other groups, as illustrated by the dot plots in Figure 8 and the cartoon in Figure 16. Also ornithine and arginine levels were decreased in the FGF21 KO as can be seen in dot plots in the Appendix (dot plots in alphabetical order by metabolite name). The Deletor pathomechanism has shifted its metabolism in such a way that keeps these three metabolites at WT levels in the Deletor FGF21 KO, whereas in the absence of FGF21 in the FGF21 KO their levels decrease. FGF21 does not affect creatine synthesis pathway.

Creatine synthesis from utilises glycine and 1C units. Guanidinoacetate is synthesised from arginine and glycine, releasing ornithine. GAA is then methylated to produce creatine. Based on the result that MTHFD2 stays upregulated in the Deletor FGF21 KO, the folate cycle is upregulated at least partly independently of FGF21. This would explain the fact that the KO does not lower creatine and GAA levels in the Deletor FGF21 KO in that sufficient methylation capacity exists to keep

their levels seemingly normal.

The short review by Ostojic (2015) speculates on the possibility that GAA could serve as an alternative energy donor when there is a lack of creatine. GAA is the direct precursor for creatine. The downregulation of creatine and guanidoacetate in the FGF21 KO suggests, that these compounds make up a further niche of energy metabolism in which FGF21 is involved as a hormone. Also, all of these compounds stay high in the Deletor FGF21 KO compared to the FGF21 KO. Could the Deletor utilise creatine phosphate more than normal as an energy source to compensate for defective oxidative phosphorylation in skeletal muscle?

The reason for prevented lowering of succinate in the Deletor FGF21 KO compared to the simple FGF21 KO mouse could lie in the OXPHOS itself. The NADH/NAD+ ratio increases when OXPHOS is defective, as reviewed in (Nunnari & Suomalainen 2012), and intuitively TCA cycle intermediates could accumulate in a similar way. The review Smeitnik et al. (2006) concludes that TCA intermediates should accumulate in tissues and body fluids and cites Esteitie et al. (2005), which shows that TCA cycle intermediates are increased in the urine of patients with mitochondrial disease. Beside being part of the TCA, succinate is also an epigenetic marker (Zhang et al. 2010). The fluctuation of its levels could influence protein activity or signalling pathways, which adds yet another level of complexity to the actions of FGF21.

Creatine and succinate have been studied as drug candidates for mitochondrial diseases in clinical trials. The study on succinate was a case report (Shoffner et al. 1989) with a positive outcome. The patient had a severe respiratory complex I deficiency, and combined succinate and coenzyme Q supplementation was administered to her. The goal of the treatment was to bypass complex I in the respiratory chain to keep oxidative phosphorylation working, and the treatment succeeded in keeping the patient stable. Creatine supplementation was tried in several studies with variable but mostly negative outcome (Pfeffer et al. 2013). The fact that levels of these compounds are sustained in the mitohondrial myopathy of Deletor mice would match with unfruitful supplementation for treatment of patients. The case report concerning treatment with succinate described above was extreme, and partially intact OXPHOS is always better than no respiration at all.

6.1.5 On metabolic inflexibility

As a compound class, the acylcarnitines constantly attract attention in the Deletor, where they accumulate in skeletal muscle (Khan et al. 2017). This study shows that their accumulation is independent of FGF21. Long chain acyl carnitines can be seen grouped together in heatmaps for the top 25 significant metabolites from ANOVA and RF in Figure 15, where they are high in

the Deletor and Deletor FGF21 KO. ANOVA and its post-hoc test judged the pattern significant as significant all but hexanoyl and decanoyl carnitine. Accumulating acylcarnitines are seen in conjunction with glucose intolerance (Koves et al. 2008). Glucose intolerance is associated with metabolic inflexibility (Muoio 2014, Goodpaster & Sparks 2017). Metabolic flexibility is the ability of an organism to adapt its metabolism according to demand, in response to e.g. energy expenditure or outside conditions.

In the Deletor, affected muscle tissues have increased glucose uptake (Nikkanen et al. 2016). PEO patients are glycolytic, and in a clinical trial their diseased muscle fibres could not tolerate a ketogenic diet (Ahola et al. 2016). In a recent review, Goodpaster & Sparks (2017) speculate on the possibility that insulin resistance might be an adaptive response to excess glucose entering and stored in muscle cells in the absence of increased energy expenditure. No abnormalities were found in glucose and insulin tolerance tests of Deletors (Tyynismaa et al. 2010). Pereira et al. (2017) state that insulin resistance is a feature seen in mitochondrial diseases and go on to show that FGF21 rescues the developing insulin resistance in OPA1-deficient mice.

Acetylcarnitine is an interesting metabolite with respect to insulin sensitivity. The carnitine acetyltransferase has been found to promote glucose tolerance and insulin action (Muoio et al. 2012). The retention of acetylcarnitine at WT levels in the Deletor FGF21 KO fits well into this picture on insulin action. The upregulation of acetylcarnitine in the disease could be an example of beneficial FGF21 action in the myopathy. Measured acetylcarnitine levels are shown in the dot plot in Figure 5.4.2.

It is an interesting point of view to consider the disease state in terms of metabolic inflexibility. Knowing that in mitochondrial diseases the OXPHOS is compromised by definition, this can be seen as superfluous, but the accumulation of acyl carnitines is found to be a stress marker for metabolic diseases, related to diabetes and obesity (Seiler et al. 2014, Noland et al. 2009). Comparing these categories of metabolic disease might highlight possible parallel lines in disease mechanism.

6.2 Technical aspects on metabolomics

6.2.1 Group classification

PCA and PLS-DA were tried out for group classification, but these methods could not separate the four genotypes. When all four groups were analysed together, the explained percentage of variation was about 30-40% in both analyses, which does not convince to draw definitive conclusions. Expecting a few principal components to be able to summarise shifts in a complex metabolic network

is probably wishful thinking, but these multivariate tools are widely used and consequently they were explored also in this study.

The highest variation in the data these methods could pinpoint did not lead further in the analysis. Because of this and because the PLS-DA model was not plausible as shown by cross validation and permutation analysis, illustrated in the Appendix in Figure A1, results from PCA and PLS-DA were disregarded in the course of analysis. As an example, the two metabolites with the highest VIP score in PLS-DA for all four genotype groups were hippurate and 4-pyridoxate. Looking at an illustration of their levels, e.g. the heatmap of all metabolites in Figure 4, these two had very high concentration values in two WT samples, and otherwise no pattern. According to both the t-test and ANOVA these two metabolites had no changes with $p < 0.05$ between any genotypes.

The difficulty in classification also has to do with the assessment of whether or not it is a meaningful question to compare these particular four genotypes all together at the same time. It is indeed more sensible to compare individual genotypes in this case, because e.g. the relationship of the FGF21 KO to the Deletor is not trivial. Nonetheless, comparison of individual genotype pairs using the multivariate methods PCA and PLS-DA also did not yield meaningful insights. Differences between genotype pairs could however be discerned using other methods tha PCA and PLS-DA, as was done e.g. for the Deletor versus the Deletor FGF21 KO in Figure 10.

As an example of unfruitful pairwise classification attempts using PCA and PLS-DA, see Figure 17 for the comparison of Deletor and Deletor FGF21 KO genotypes. PCA could not separate the groups, although it points out N5-formyl-THF in the loadings plot. The supervised method PLS-DA neatly shows separation, but the only metabolite it uses is again N5-formyl-THF. Looking at all measured folate intermediates in this data set as depicted in Figure A3 in the Appendix, this particular folate metabolite does not conform with the levels of other folates in this metabolomic set. It also had a missing value that was imputed. No conclusions concerning the folate cycle could be drawn based on this result.

6.2.2 Previouly published results not replicated exactly

Homogeneity of the data set hampered with both analysis using multivariate methods and replication of previous results. Most of the metabolite level alterations previously reported in Deletors (Nikkanen et al. 2016) were not reproduced in the Deletors of this study, as is shown in Table 6. Here, all metabolites reported as significantly changed (all of them upregulated, with $p < 0.05$ based on Student's t-test) were listed and respective results from this study were compared to them. With a large set of measurements, it is to be expected that variance is observed (Bender & Lange 2001),

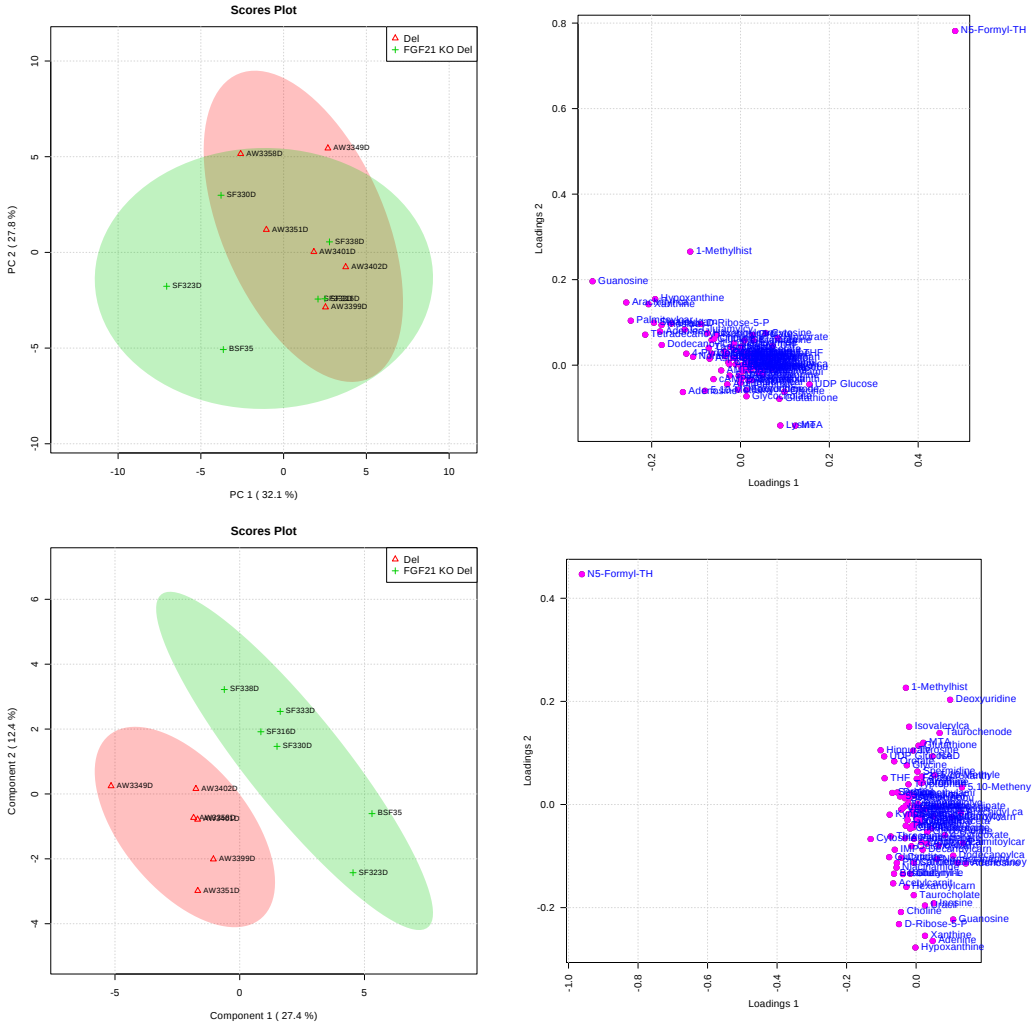


Figure 17: Comparison between the Deletor and Deletor FGF21 KO using multivariate methods of group classification. Above: A principal component analysis (PCA) scores plot of the first two principal components and the corresponding loadings plot. Below: A partial least squares discriminant analysis (PLS-DA) scores plot of the first two components and the corresponding loadings plot.

especially of metabolites which have quicly fluctuating levels. Still, the upregulation of serine and glycine which are essential for the Deletor phenotype were not seen here.

Instead, e.g. in all of the pairwise comparisons using PCA and PLS-DA, the main differences between mouse groups in this study as pointed out by loadings plots were made by metabolites with concentration values that were spread out with low accuracy, e.g. hippurate, 4-pyridoxate, N5-formyl-THF, or guanosine. Also the OOB error rate of the RF classification in this study was 0.5, which is very high compared to the MetaboAnalyst example data error rate in RF, which was 0.154 (Xia & Wishart 2016).

Thankfully, e.g. ANOVA listed significant changes in 32 metabolites so the issue is not that there was not any variation in the data. Insight into the changes had to be sought elsewhere than from PCA, PLS-DA, or direct comparison with data from Nikkanen et al. (2016).

6.2.3 Missing values and outliers

The fact that missing value imputation can have a large effect on statistical analysis (Armitage et al. 2015), was seen in this study. The four missing values for the sample AW3400D- were at first imputed by a small positive value: half of the minimum detected value. This method of imputation is designed for the case where concentrations cannot be measured because they are below the lowest instrumental detection limit, as is often the case for metabolites. Clearly this was not the case for these particular metabolites, because the imputed low values stood out to such extent, that the sample looked like an extreme outlier on PCA plots.

The error was noticed when comparing the global heatmap with the table of original concentration values. Changing the imputation method to replacement by column mean clarified the at first problematic clustering, and subsequent analysis did not implicate any further difficulty with this particular sample.

The RF algorithm can also be used for outlier detection. Possible outliers suggested by RF are depicted in in Figure 18. Highest among these was the sample SF324D-, which attracted attention as a possible outlier also in other experiments done on the QF tissue of that individual mouse. Because several samples had similar scores, and none of them stood out, no outliers were removed based on the RF analysis or for any other reason, except in t-tests for comparability with Nikkanen et al. (2016) as described in the Methods Section.

Pathway	Reported as upregulated	replicated with p<0.05?	FC
Folate cycle	5-methyl-THF	yes	1.42
	5-methylene-THF	-	-
	formate	-	-
	glycine	no	1.47
	serine	no	1.19
	THF	no	1.01
Amino acids	alanine	no	1.01
	asparagine	no	1.31
	aspartate	no	1.26
	glutamic acid	yes	1.90
	glutamine	no	1.35
	homoserine	-	-
	hydroxyproline	-	-
	isoleucine	no	1.12
	leucine	no	1.13
	phenylalanine	no	1.09
	proline	no	1.27
	threonine	yes	1.27
	valine	no	1.19
Methyl cycle	guanidinoacetic acid	no	1.12
	methionine	no	1.17
Purine synthesis	hypoxanthine	no	0.86
	5-hydroxyindole-3-acetic acid	-	-
	betaine	yes	1.60
	choline	yes	1.82
	creatine	no	1.01
	creatinine	no	1.14
	GABA	-	-
	glutathione	no	0.90
	glyceraldehyde	no	1.10
	neopterin	no	1.05
	niacinamide	yes	1.31
	phosphoethanolamine	yes	1.46
	taurine	no	1.13

Table 6: All significantly changed metabolites in Deletor QF from Nikkanen et al. (2016) and whether the results are replicated in this study. The referred p-values were determined by t-test. Metabolites which were not measured in this data set are also listed.

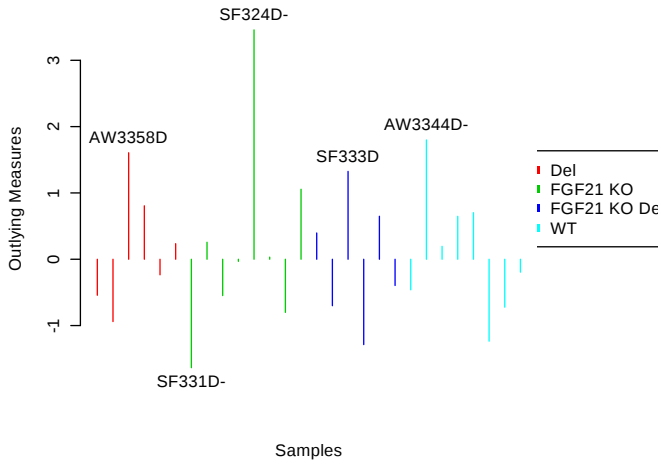


Figure 18: Outlier prediction in random forests (RF). None of the indicated samples was ultimately removed as an outlier.

6.2.4 Further correlation analysis

The PatternHunter correlation analysis in Metaboanalyst can be used for searching for metabolites with specific concentration patterns, where patterns are indicated by choosing arbitrary integer numbers to describe concentration differences. In the analysis for this study, the PatternHunter did not reveal anything about the data that had not been noticed with other methods already. It is however a convenient way of comparison that yields insightful illustrations, and I recommend it for use at other instances. Below is a brief description of its use in this study.

Because almost all of the significantly changed metabolites in the Deletor tend to be upregulated, patterns were chosen to be searched for in which a metabolite is high in the Deletor and low in other groups. The Pearson R was used for measuring correlation. The pattern numbering was chosen so that it would be simple, but still 3 was used instead of 2 to emphasise the upregulation in the Deletor. In the analysis the groups were arranged in the order Del - KO - DelKO - WT. The pattern 3-1-3-1 reveals, which upregulated metabolites of the Deletor are induced independently of FGF21, since they stay up in the DelKO mice. The pattern 3-1-1-1 again reveals, which metabolites are dependent on FGF21 induction. The results can be seen in Figure 19.

The results show, that there are fairly large groups of metabolites correlating both to the pattern of "upregulated in the Deletor and rescued by the KO of FGF21" and "upregulated in the Deletor and not affected byt the KO of FGF21". The Pearson R:s were not extremely high, but these general trends had already been seen in the heatmaps and in the analysis of individual metabolites. As

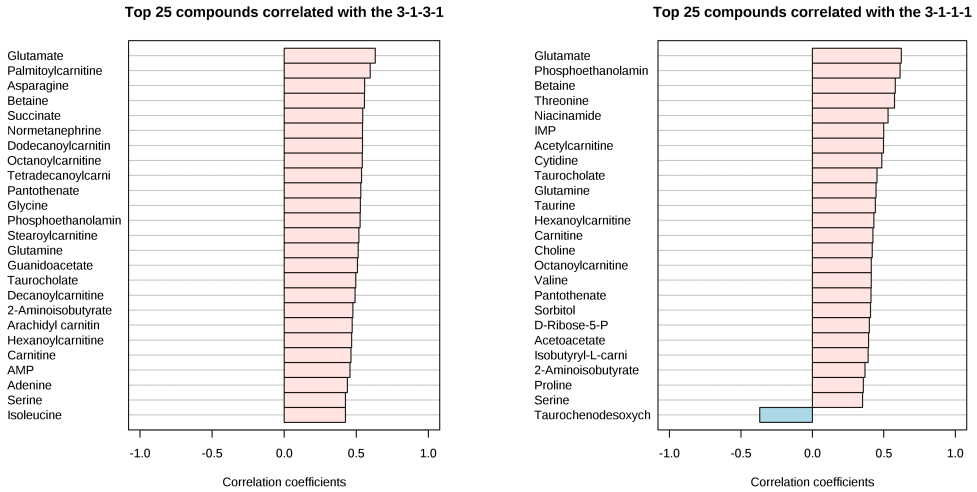


Figure 19: PatternHunter correlation search for patterns in metabolite levels across groups. The genotype groups were arranged in the order Del - KO - DelKO - WT. The pattern 3-1-3-1 (left) shows, which upregulated metabolites of the Deletor are induced independently of FGF21, since they stay up in the DelKO mice. The pattern 3-1-1-1 (right) shows which metabolites are dependent on FGF21 induction. The Pearson R is used for measuring correlation.

the Pearson R only reaches values of about 0.5 in this analysis, many of the metabolites are shared between both patterns. The upregulation in the Deletor then is the main common denominator for these other metabolites.

6.3 Future considerations and dreams

6.3.1 Biology

A simple next step that directly aligns with the concept of the study in this thesis would be to measure this same set of metabolites in the heart tissue of the studied mice. The Deletor heart was shown to exhibit extensive metabolic alterations that partly overlap with the QF (Nikkanen et al. 2016), and looking into the heart would doubtlessly complement the results obtained so far from QF.

Another perspective would be to look into the compartmentalisation of metabolite levels and metabolic changes. Chen et al. (2016) developed a method of isolating mitochondria and performing a metabolomic experiment on this cellular compartment alone. It would be very interesting to know something about local concentrations of metabolites in individual cellular compartments. This help to understand e.g. cell signalling. If a metabolite acts as a signalling molecule, its concentration at

a specific location matters. Also pathway imbalances would be revealed in more detail, deciphering physiology and pathophysiology.

6.3.2 Experimental setup

To address the homogeneity of the metabolomic data, basic refinement of the experimental setup might be beneficial. The number of biological replicates could be optimised. Power calculations could be performed to estimate the required number of samples for seeing differences more clearly. In this study, the number of samples per group varied between 6–8. Also stricter standardisation of sample collection might be useful to look into, since small molecular metabolite levels fluctuate rapidly (personal communication, Vidya Velagapudi). Both of these approaches require some restructuring of the routines with which these experiments are performed.

In a bigger picture, standardisation of experiments makes comparison between studies possible. International organisations like the Metabolomics Standards Initiative (MSI) or the European Commission funded COSMOS Project are concentrating efforts on the standardisation of metabolomics in general. When logging sufficient metadata and adhering to good practises, metabolomics data can be admitted for upload into databases like MetaboLights (Haug et al. 2012). Like other omics data, e.g. from DNA sequencing or microarray experiments, the data can then be reanalysed and used for meta-analysis or other large studies conducted by third parties.

6.3.3 Bioinformatics

MetaboAnalyst offers even more tools, not to speak of computational possibilities beyond the MetaboAnalyst environment, that were not used in this thesis but could be used in further studies. To pick one possibility, gene set enrichment analysis (GSEA) is a tool for microarray data interpretation developed by Subramanian et al. (2005). It was modified for use in metabolomics by Xia & Wishart (2010) and integrated into MetaboAnalyst with the name metabolite set enrichment analysis (MSEA). The idea behind enrichment analysis is to be able to detect widespread changes in the searched sets while variation in individual features might be small in scale.

Enrichment analysis avoids the confusion arising from multiple testing and its adjustments for it. Instead of obtaining lists of unreliable significances or hardly any significant changes at all, it MSEA is designed to detect subtle changes, as long as these are widespread. Analogous to SAM, it is an analytical tool tailored for the characteristics of 'omics data. Both were originally invented to counter the needs of transcriptomic analysis, and were later modified for use in the younger

metabolomic field (Xia & Wishart 2016).

A possibility for improving future studies might thus be to define metabolite sets of interest that have been established to be of particular interest for the research topic, or that have been hypothetically indicated as relevant. The predefined sets based on the KEGG database that MSEA in MetaboAnalyst uses are often too general (e.g. "protein synthesis") or too narrow (only one metabolite hit fits the set) to yield meaningful insight, but self-defined sets could be more helpful. Measuring a larger set of metabolites would also help in producing more insightful results, but development of the targeted metabolomics experiment is a whole field of its own.

7 Conclusions

As an applicable biomarker FGF21 has been established a prominent feature of the pathology in mitochondrial myopathies. When knowing nothing more about its role in these diseases except for its robust upregulation, there are two hypothetical options for its actions. Either FGF21 is a protective measure that the body summons to counteract the disease, or it is part of the disease driving pathophysiology.

The FGF21 KO project using Deletor mice investigates this. As part of the project, the study presented in this thesis establishes that FGF21 contributes to the muscle pathology. The results from the metabolomic analysis and the measured enzyme expression levels show, that essential features of the metabolic changes of the diseased muscle are mediated by FGF21. It increases the flux of carbon units into one carbon metabolism via several routes – both serine de novo synthesis and individual alternative carbon donors – and it increases enzyme expression of the transsulfuration pathway. However, the results do not exlude the possibility that these changes could be protective for the organism.

Additionally, shifts in the levels of metabolites in energy metabolism in KO mice with WT background indicate, that as a hormone FGF21 has in normal physiology an influence on energy metabolism beyond the regulation of lipolysis in adipose tissue. Further metabolomic inquiries and replication of the findings of this study will give more detailed knowledge about the actions of FGF21 in the future.

References

Ahola-Erkkilä, S., Carroll, C., Peltola-Mjösund, K., Tulkki, V., Mattila, I., Seppänen-Laakso, T., Oresic, M., Tyynismaa, H. & Suomalainen, A. (2010), 'Ketogenic diet slows down mitochondrial myopathy progression in mice', *Hum Mol Genet* **19**, 1974–1984. DOI: 10.1093/hmg/ddq076.

Ahola, S., Auranen, M., Isohanni, P., Niemisalo, S., Urho, N., Buzkova, J., Velagapudi, V., Lundbom, N., Hakkarainen, A., Muurinen, T., Piirilä, P., Pietiläinen, K. H. & Suomalainen, A. (2016), 'Modified atkins diet induces subacute selective ragged-red-fiber lysis in mitochondrial myopathy patients', *EMBO Mol Med* **8**, 1234–1248. DOI: 10.15252/emmm.201606592.

Armitage, E., Godzien, J., Alonso-Herranz, V., Lopez-Gonzalvez, A. & Barbas, C. (2015), 'Missing value imputation strategies for metabolomics data', *Electrophoresis* **36**, 3050–3060. DOI: 10.1002/elps.201500352.

Bao, X. R., Ong, S.-E., Goldberger, O., Peng, J., Sharma, R., Thompson, D. A., Vafai, S. B., Cox, A. G., Marutani, E., Ichinose, F., Goessling, W., Regev, A., Carr, S. A., Clish, C. B. & Mootha, V. K. (2016), 'Mitochondrial dysfunction remodels one-carbon metabolism in human cells', *Elife* **5:e10575**. DOI: 10.7554/eLife.10575.

Bender, R. & Lange, S. (2001), 'Adjusting for multiple testing - when and how?', *J Clin Epidemiol* **54**, 343–349.

Breiman, L. (2001), 'Random forests', *Mach Learn* **45**, 5–32.

Chadeau-Hyam, M., Campanella, G., Jombart, T., Bottolo, L., Portengen, L., Vineis, P., Liquet, B. & Vermeulen, R. C. (2013), 'Deciphering the complex: Methodological overview of statistical models to derive omics-based biomarkers', *Environ Mol Mutagen* **54**, 542–557. DOI: 10.1002/em.21797.

Chang, W. (2013), *R Graphics Cookbook*, O'Reilly.

Chen, W. W., Freinkman, E., Wang, T., Birsoy, K. & Sabatini, D. M. (2016), 'Absolute quantification of matrix metabolites reveals the dynamics of mitochondrial metabolism', *Cell* **166**, 1324–1337. DOI: 10.1016/j.cell.2016.07.040.

Copeland, W. (2008), 'Inherited mitochondrial diseases of DNA replication', *Annu Rev Med* **59**, 131–146.

Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. (2006), 'Method to account for dilution of complex biological mixtures. application in 1H NMR metabonomics', *Anal Chem* **78**, 4281–4290. DOI 10.1021/ac051632c.

Ducker, G. S. & Rabinowitz, J. D. (2017), 'One-carbon metabolism in health and disease', *Cell Metab* **25**, 27–43. DOI: 10.1016/j.cmet.2016.08.009.

Ebbels, T. M. D. & Iorio, M. D. (2011*a*), Statistical data analysis in metabolomics, *in* M. P. H. Stumpf, D. J. Balding & M. Girolami, eds, 'Handbook of Statistical Systems Biology', 1 edn, John Wiley and Sons, Ltd, chapter 8, pp. 163–180. ISBN 978-0-470-71086-9.

Ebbels, T. M. D. & Iorio, M. D. (2011*b*), Statistical data analysis in metabolomics, *in* M. P. Stumpf, D. J. Balding & M. Girolami, eds, 'Handbook of Statistical Systems Biology', John Wiley et Sons, Ltd, chapter 8, pp. 163–180. ISBN 978-0-470-71086-9.

Edgar, A. J. (2002), 'The human l-threonine 3-dehydrogenase gene is an expressed pseudogene', *BMC Genet* **3**. DOI: 10.1186/1471-2156-3-18.

Esteitie, N., Hinttala, R., Wibom, R., Nilsson, H., Hance, N., Naess, K., Teär-Fahnehjelm, K., von Döbeln, U., Majamaa, K. & Larsson, N.-G. (2005), 'Secondary metabolic effects in complex I deficiency', *Ann Neurol* **58**, 544–552. DOI: 10.1002/ana.20570.

Fischer, F. M. & Marator-Flier, E. (2016), 'Understanding the physiology of FGF21', *Annu Rev Physiol* **78**, 223–241. DOI: 10.1146/annurev-physiol-021115-105339.

Goffart, S., Cooper, H., Tyynismaa, H., Wanrooij, S., Suomalainen, A. & Spelbrink, J. (2009), 'Twinkle mutations associated with autosomal dominant progressive external ophthalmoplegia lead to impaired helicase function and in vivo mtDNA replication stalling', *Hum Mol Genet* **18**, 328–340. DOI: 10.1093/hmg/ddn359.

Goodpaster, B. H. & Sparks, L. M. (2017), 'Metabolic flexibility in health and disease', *Cell Metab* **25**, 1027–1037. DOI: 10.1016/j.cmet.2017.04.015.

Gorman, G. S., Chinnery, P. F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D. R., Zeviani, M. & Turnbull, D. M. (2016), 'Mitochondrial diseases', *Nat Rev Disease Primers* **2**, 1–22. DOI: 10.1038/nrdp.2016.80.

Gowda, H., Ivanisevic, J., Johnson, C., Kurczy, M., Benton, P., Rinehart, D., Nguyen, T., Ray, J., Kuehl, J., Arevalo, B., Westenskow, P., Wang, J., Arkin, A., Deutschbauer, A., Patti, G. & Siuzdak, G. (2014), 'Interactive XCMS online: Simplifying advanced metabolomic data processing and subsequent statistical analyses', *Anal Chem* . DOI: 10.1021/ac500734c.

Haug, K., Salek, R. M., Conesa, P., Hastings, J., de Matos, P., Rijnbeek, M., Mahendraker, T., Williams, M., Neumann, S., Rocca-Serra, P., Maguire, E., Gonzalez-Beltra, A., Sansone, S.-A., Griffin, J. L. & Steinbeck, C. (2012), 'Metabolights–an open-access general-purpose repository

for metabolomics studies and associated meta-data', *Nucleic Acids Res* **41**, D781–D786. DOI: 10.1093/nar/gks1004.

Inagaki, T., Dutchak, P., Zhao, G., Ding, X., Gautron, L., Parameswara, V., Li, Y., Goetz, R., Mohammadi, M., Esser, V., Elmquist, J., Gerard, R., Burgess, S., Hammer, R., Mangelsdorf, D. & Kliewer, S. (2007), 'Endocrine regulation of the fasting response by PPARα-mediated induction of fibroblast growth factor 21', *Cell Metab* **5**, 415–425. DOI: 10.1016/j.cmet.2007.05.003.

Johnson, C. H., Ivanisevic, J. & Siuzdak, G. (2016), 'Metabolomics: beyond biomarkers and towards mechanisms', *Nat Rev Mol Cell Bio* **17**, 451–459.

Kalogeropoulou, A. (2011), Pre-processing and analysis of high-dimensional plant metabolomic data, Master's thesis, University of East Anglia, Norwich, England.

Khan, N. A., Nikkanen, J., Yatsuga, S., Jackson, C., Wang, L., Pradhan, S., Kivelä, R., Pessia, A., Velagapudi, V. & Suomalainen, A. (2017), 'mTORC1 regulates mitochondrial integrated stress response and mitochondrial myopathy progression', *Cell Metab* **26**, 419–428. DOI: 10.1016/j.cmet.2017.07.007.

Kharitonenkov, A. & DiMarchi, R. (2017), 'Fibroblast growth factor 21 night watch: advances and uncertainties in the field', *J Intern Med* **281**, 233–246. DOI: 10.1111/joim.12580.

Kharitonenkov, A., Shiyanova, T., Koester, A., Ford, A., Micanovic, R., Galbreath, E., Sandusky, G., Hammond, L., Moyers, J., Owens, R., Gromada, J., Brozinick, J., Hawkins, E., Wroblewski, V., Li, D.-S., Mehrbod, F., Jaskunas, S. R. & Shanafelt, A. (2005), 'FGF-21 as a novel metabolic regulator', *J Clin Invest* **115**, 1627–1635.

Kharitonenkov, A., Wroblewski, V., Koester, A., Chen, Y.-F., Clutinger, C., Tigno, X., Hansen, B., Shanafelt, A. & Etgen, G. (2006), 'The metabolic state of diabetic monkeys is regulated by fibroblast growth factor-21', *Endocrinology* **148**, 774–781. DOI: 10.1210/en.2006-1168.

Kilberg, M., Pan, Y.-X., Chen, H. & Leung-Pineda, V. (2005), 'Nutritional control of gene expression: how mammalian cells respond to amino acid limitation', *Annu Rev Nutr* **25**, 59–85. DOI: 10.1146/annurev.nutr.24.012003.132145.

Koskinen, T., Santavuori, P., Sainio, K., Lappi, M., Kallio, A.-K. & Pihko, H. (1994), 'Infantile onset spinocerebellar ataxia with sensory neuropathy: a new inherited disease', *J Neurol Sci* **121**, 50–56. DOI: 10.1016/0022-510X(94)90156-2.

Koves, T. R., Ussher, J. R., Noland, R. C., Slentz, D., Mosedale, M., Ilkayeva, O., Bain, J., Stevens, R., Dyck, J. R., Newgard, C. B., Lopashuk, G. D. & Muoio, D. M. (2008), 'Mitochondrial overload

and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance', *Cell Metab* **7**, 45–57. DOI: 10.1016/j.cmet.2007.10.013.

Lallemand, Y., Luria, V., Haffner-Krausz, R. & Lonai, P. (1998), 'Maternally expressed PKG-Cre transgene as a tool for early und uniform activation of the Cre site-specific recombinase', *Transgenic Res* pp. 105–112.

Lehtonen, J. M., Forsström, S., Bottani, E., Viscomi, C., Baris, O. R., Isoniemi, H., Höckerstedt, K., Österlund, P., Hurme, M., Jylhävä, J., Leppä, S., Markkula, R., Heliö, T., Mombelli, G., Uusimaa, J., Laaksonen, R., Laaksovirta, H., Auranen, M., Zeviani, M., Smeitnik, J., Wiesner, R. J., Nakada, K., Isohanni, P. & Suomalainen, A. (2016), 'FGF21 is a biomarker for mitochondrial translation and mtDNA maintenance disorders', *Neurology* **87**, 2290–2299.

Liu, X. & Locasale, J. W. (2017), 'Metabolomics: A Primer', *Trends Biochem Sci* **1**, 1. DOI: 10.1016/j.tibs.2017.01.004.

Maruyama, R., Shimizu, M., Li, J., Inoue, J. & Sato, R. (2016), 'Fibroblast growth factor 21 induction by activating transcription factor 4 is regulated through three amino acid response elements in its promoter region', *Biosci Biotech Bioch* **5**, 929–934. DOI: 10.1080/09168451.2015.1135045.

Mary-Huard, T. & Robin, S. (2011), Introduction to statistical methods for complex systems, *in* M. P. Stumpf, D. J. Balding & M. Girolami, eds, 'Handbook of Statistical Systems Biology', John Wiley et Sons, Ltd, chapter 2, pp. 15–38. ISBN 978-0-470-71086-9.

Miller, L. (2010), 'Analyzing gels and western blots with ImageJ'. read on 27.11.2016.
URL: *http://lukemiller.org/index.php/2010/11/analyzing-gels-and-western-blots-with-image-j/*

Minard, A. Y., Tan, S.-X., Yang, P., Fazakerley, D. J., Domanova, W., Parker, B. L., Humphrey, S. J., Jothi, R., Stöckli, J. & James, D. E. (2016), 'mTORC1 is a major regulatory node in the FGF21 signaling network in adipocytes', *Cell Rep* **17**, 29–37. DOI: 10.1016/j.celrep.2016.08.086.

Muoio, D. M. (2014), 'Metabolic inflexibility: When mitochondrial indecision leads to metabolic gridlock', *Cell* **159**, 1253–1353. DOI: 10.1016/j.cell.2014.11.034.

Muoio, D. M., Noland, R. C., Kovalik, J.-P., Seiler, S. E., Davies, M. N., DeBalsi, K. L., Ilkayeva, O. R., Stevens, R. D., Kheterpal, I., Zhang, J., Covington, J. D., Bajpeyi, S., Ravussin, E., Kraus, W., Koves, T. R. & Mynatt, R. L. (2012), 'Muscle-specific deletion of carnitine acetyltransferase compromises glucose tolerance and metabolic flexibility', *Cell Metab* **15**, 764–778. DOI: 10.1016/j.cmet.2012.04.005.

Nikkanen, J., Forsström, S., Euro, L., Paetau, I., Kohnz, R. A., Wang, L., Chilov, D., Viinamäki, J., Roivainen, A., Marjamäki, P., Liljenbäck, H., Ahola, S., Buzkova, J., Terzioglu, M., Khan, N. A.,

Pirnes-Karhu, S., Paetau, A., Lönnqvist, T., Sajantila, A., Isohanni, P., Tyynismaa, H., Nomura, D. K., Battersby, B. J., Velagapudi, V., Carroll, C. J. & Suomalainen, A. (2016), 'Mitochondrial DNA replication defects disturb cellular dNTP pools and remodel one-carbon metabolism', *Cell Metab* **23**, 635–648. DOI 10.1016/j.cmet.2016.01.019.

Nilsson, R., Jain, M., Madhusudhan, N., Sheppard, N. G., Strittmatter, L., Kampf, C., Huang, J., Asplund, A. & Mootha, V. K. (2014), 'Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer', *Nat Commun* **5**. DOI: 10.1038/ncomms4128.

Nishimura, T., Nakatake, Y., Konishi, M. & Itoh, N. (2000), 'Identification of a novel FGF, FGF-21, preferentially expressed in the liver', *Biochim Biophys Acta* **1492**, 203–206.

Noland, R. C., Koves, T. R., Seiler, S. E., Lum, H., Lust, R. M., Ilkayeva, O., Stevens, R. D., Hegardt, F. G. & Muoio, D. M. (2009), 'Carnitine insufficiency caused by aging and overnutrition compromises mitochondrial performance and metabolic control', *J Biol Chem* **284**, 22840–22852. DOI: 10.1074/jbc.M109.032888.

Nunnari, J. & Suomalainen, A. (2012), 'Mitochondria: In sickness and in health', *Cell* **148**, 1145–1159. DOI: 10.1016/j.cell.2012.02.035.

Ostojic, S. M. (2015), 'Cellular bioenergetics of guanidinoacetic acid: the role of mitochondria', *J Bioenerg Biomembr* **47**, 369–372. DOI: 10.1007/s10863-015-9619-7.

Pereira, R. O., Tadinada, S. M., Zasadny, F. M., Oliveira, K. J., Pires, K. M. P., Olvera, A., Jeffers, J., Souvenir, R., Mcglaufin, R., Seei, A., Funari, T., Sesaki, H., Potthoff, M. J. & Adams, C. M. (2017), 'OPA1 deficiency promotes secretion of FGF21 from muscle that prevents obesity and insulin resistance', *EMBO J* **36**, 2126–2145. DOI: 10.15252/embj.201696179.

Pfeffer, G., Horvath, R., Klopstock, T., Mootha, V. K., Suomalainen, A., Koene, S., Hirano, M., Zeviani, M., Bindoff, L. A., Yu-Wai-Man, P., Hanna, M., Carelli, V., McFarland, R., Majamaa, K., Turnbull, D. M., Smeitink, J. & Chinnery, P. F. (2013), 'New treatments for mitochondrial disease–no time to drop our standards', *Nat Rev Neurol* **9**, 474–481. DOI: 10.1038/nrneurol.2013.129.

Potthoff, M. (2016), 'A new frontier in FGF21 biology', *Nat Rev Endocrinol* . DOI: 10.1038/nrendo.2016.206.

Potthoff, M., Inagaki, T., Satapati, S., Ding, X., He, T., Goetz, R., Mohammadi, M., Finck, B., Mangelsdorf, D., Kliewer, S. & Burgess, S. (2009), 'FGF21 induces PGC-1α and regulates carbohydrate and fatty acid metabolism during the adaptive starvation response', *PNAS* **106**, 10853–10858. DOI: 10.1073 pnas.0904187106.

Ressom, H. W., Varghese, R. S., Zhang, Z., Xuan, J. & Clarke, R. (2008), 'Classification algorithms for phenotype prediction in genomics and proteomics', *Front Biosci* **13**, 691–708. DOI: 10.2741/2712.

Salminen, A., Kaarniranta, K. & Kauppinen, A. (2017), 'Integrated stress response stimulates FGF21 expression: Systemic enhancer of longevity', *Cell Signal* **40**, 10–21. DOI: 10.1016/j.cellsig.2017.08.009.

Schmittgen, T. D. & Livak, K. J. (2008), 'Analyzing real-time PCR data by the comparative Ct method', *Nat Protoc* **3**, 1101–1108. DOI: 10.1038/nprot.2008.73.

Seiler, S. E., Martin, O. J., Noland, R. C., Slentz, D. H., DeBalsi, K. L., Ilkayeva, O. R., An, J., Newgard, C. B., Koves, T. R. & Muoio, D. M. (2014), 'Obesity and lipid stress inhibit carnitine acetyltransferase acivity', *J Lipid Res* **55**, 635–645. DOI: 10.1194/jlr.M043448.

Shoffner, J. M., Lott, M. T., Voljavec, A. S., Soueidan, S. A., Costigan, D. A. & Wallace, D. C. (1989), 'Spontaneous Kearns-Sayre/chronic external opththalmoplegia plus syndrome associated with a mitochondrial DNA deletion: A slip-replication model and metabolic therapy', *Proc Natl Acad Sci USA* **86**, 7952–7956.

Smeitnik, J. A., Zeviani, M., Turnbull, D. M. & Jacobs, H. T. (2006), 'Mitochondrial medicine: A metabolic prespective on the pathology of oxidative phosphorylation disorders', *Cell Metab* **3**, 9–13. DOI: 10.1016/j.cmet.2005.12.001.

Sonoda, J., Chen, M. Z. & Baruch, A. (2017), 'FGF21-receptor agonists: an emerging therapeutic class for obesity-related diseases', *Horm Mol Biol Clin Investig* **30**. DOI: 10.1515/hmbci-2017-0002.

Spelbrink, J. N., Li, F.-Y., Tiranti, V., Nikali, K., Yuan, Q.-P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.-M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H. T., Zeviani, M. & Larsson, C. (2001), 'Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localised in mitochondria', *Nat Genet* **28**, 223–231. DOI: 10.1038/90058.

Srivastava, S. & Moraes, C. (2005), 'Double-strand breaks of mouse muscle mtdna promote large deletions similar to multiple mtDNA deletions in humans', *Hum Mol Genet* **14**, 893–902. DOI: 10.1093/hmg/ddi082.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjeed, S., Eberta, B. L., Gillette, M. A., Paulovichg, A., Pomeroyh, S. L., Goluba, T. R., Landera, E. S. & Mesirov, J. P. (2005), 'Gene

set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *PNAS* **102**, 15545–15550. DOI: 10.1073 pnas.0506580102.

Suomalainen, A. (2011b), 'Therapy for mitochondrial disorders: Little proof, high research activity, some promise', *Semin Fetal Neonat M* **16**, 236–240. DOI: 10.1016/j.siny.2011.05.003.

Suomalainen, A., Elo, J. M., Pietiläinen, K. H., Hakonen, A. H., Sevastianova, K., Korpela, M., Isohanni, P., Marjavaara, S. K., Tyni, T., Kiuru-Enari, S., Pihko, H., Darin, N., Ounap, K., Kluijtmans, L. A. J., Paetau, A., Buzkova, J., Bindoff, L. A., Annunen-Rasila, J., Uusimaa, J., Rissanen, A., Yki-Järvinen, H., Hirano, M., Tulinius, M., Smeitnik, J. & Tyynismaa, H. (2011a), 'FGF-21 as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies: a diagnostic study', *Lancet* **10**, 806–818. DOI 10.1016/S1474- 4422(11)70155-7.

Suomalainen, A., Majander, A., Haltia, M., Somer, H., Lönnqvist, J., Savontaus, M.-L. & Peltonen, L. (1992), 'Multiple deletions of mitochondrial DNA in several tissues of a patient with severe retarded depression and familial progressive external ophthalmoplegia', *J Clin Invest* **90**, 61–66. DOI: 10.1172/JCI115856.

Tautenhahn, R., Patti, G., Rinehart, D. & Siuzdak, G. (2012), 'XCMS online: A web-based platform to process untargeted metabolomic data', *Anal Chem* **84**, 5035–5039. DOI: 10.1021/ac300698c.

Thorburn, D. R. (2004), 'Mitochondrial disorders: Prevalence, myths and advances', *J Inherit Metab Dis* **27**, 349–362. DOI: 10.1023/B:BOLI.0000031098.41409.55.

Tusher, V. G., Tibshirani, R. & Chu, G. (2001), 'Significance analysis of microarrays applied to the ionizing radiation response', *PNAS* **98**, 5116–5121. DOI 10.1073 pnas.091062498.

Tyynismaa, H., Carroll, C. J., Raimundo, N., Ahola-Erkkilä, S., Wenz, T., Ruhanen, H., Guse, K., Hemminki, A., Peltola-Mjosund, K. E., Tulkki, V., Oresic, M., Moraes, C. T., Pietiläinen, K., Hovatta, I. & Suomalainen, A. (2010), 'Mitochondrial myopathy induces a starvation-like response', *Hum Mol Genet* **19**, 3948–3958. DOI 10.1093/hmg/ddq310.

Tyynismaa, H., Mjosund, K. P., Wanrooij, S., Lappalainen, I., Ylikallio, E., Jalanko, A., Spelbrink, J. N., Paetau, A. & Suomalainen, A. (2005), 'Mutant mitochondrial helicase Twinkle causes multiple mtDNA deletions and a late-onset mitochondrial disease in mice', *P Natl Acad Sci USA* **102**, 17687–17692. DOI 10.1073 pnas.0505551102.

Tyynismaa, H. & Suomalainen, A. (2009), 'Mouse models of mitochondrial DNA defects and their relevance for human disease', *EMBO Rep* **10**, 137–144. DOI: 10.1038/embor.2008.242.

Wold, S., Sjöström, M. & Eriksson, L. (2001), 'PLS-regression: a basic tool of chemometrics', *Chemometr Intell Lab* **58**, 109–130.

Worley, B. & Powers, R. (2013), 'Multivariate analysis in metabolomics', *Current Metabolomics* **1**, 92–107.

Xi, B., Gu, H., Baniasadi, H. & Raftery, D. (2014), 'Statistical analysis and modeling of mass spectrometry-based metabolomics data', *Methods Mol Biol* **1198**, 333–353. DOI: 10.1007/978-1-4939-1258-2-22.

Xia, J., Psychogios, N., Young, N. & Wishart, D. S. (2009), 'MetaboAnalyst: a web server for metabolomic data analysis and interpretation', *Nucleic Acids Res* **37**, W652–W660. DOI 10.1093/nar/gkp356.

Xia, J. & Wishart, D. S. (2010), 'MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data', *Nucleic Acids Res* **38**, W71–W77. DOI: 10.1093/nar/gkq329.

Xia, J. & Wishart, D. S. (2016), 'Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis', *Curr Protoc Bioinform* **55**, 14.10.1–14.10.91. DOI 10.1002/cpbi.11.

Ylikallio, E. & Suomalainen, A. (2012), 'Mechanisms of mitochondrial diseases', *Ann Med* **44**, 41–59. DOI 10.3109/07853890.2011.598547.

Zhang, Z., Tan, M., Xie, Z., Dai, L., Chen, Y. & Zhao, Y. (2010), 'Identification of lysine succinylation as a new post-translational modification', *Nat Chem Biol* **7**, 58–66. DOI: 10.1038/nchembio.495.

Zong, W.-X., Rabinowitz, J. D. & White, E. (2016), 'Mitochondria and cancer', *Mol Cell* **61**, 667–676. DOI: 10.1016/j.molcel.2016.02.011.

A Appendix

N:o	Metabolite	KEGG	HMDB
1	1-Methylhistamine	C05127	HMDB00898
2	2-Aminoisobutyrate	C03665	HMDB01906
3	3-Hydroxanthranilate	C00632	HMDB01476
4	4-Pyridoxate	C00847	HMDB00017
5	5-Methyl-THF	C00440	HMDB01396
6	5-10-Methenyl-THF	C00445	HMDB01354
7	Acetoacetate	C00164	HMDB00060
8	Acetylcarnitine	C02571	HMDB00201
9	Adenine	C00147	HMDB00034
10	Adenosine	C00212	HMDB00050
11	Alanine	C00041	HMDB00161
12	Allantoin	C01551	HMDB00462
13	AMP	C00020	HMDB00045
14	Arachidyl carnitine	NA	HMDB06460
15	Arginine	C00062	HMDB00517
16	Asparagine	C00152	HMDB00168
17	Aspartate	C00049	HMDB00191
18	Betaine	C00719	HMDB00043
19	Carnitine	C00318	HMDB00062
20	Chenodeoxycholate	C02528	HMDB00518
21	Choline	C00114	HMDB00097
22	Citrulline	C00327	HMDB00904
23	Creatine	C00300	HMDB00064
24	Creatinine	C00791	HMDB00562
25	Cysteine	C00097	HMDB00574
26	Cytidine	C00475	HMDB00089
27	Cytosine	C00380	HMDB00630
28	D-Ribose-5-P	C00117	HMDB01548
29	Decanoylcarnitine	NA	HMDB00651
30	Deoxycytidine	C00881	HMDB00014
31	Deoxyuridine	C00526	HMDB00012
32	Dimethylglycine	C01026	HMDB00092
33	Dodecanoylcarnitine	NA	HMDB02250
34	Folate	C00504	HMDB00121
35	G-Glutamylcysteine	C00669	HMDB01049
36	Glutamate	C00025	HMDB00148
37	Glutamine	C00064	HMDB00641
38	Glutathione	C00127	HMDB00125
39	Glyceraldehyde	C02154	HMDB01051
40	Glycine	C00037	HMDB00123

41	Glycocholate	C01921	HMDB00138
42	Guanidoacetate	C00581	HMDB00128
43	Guanosine	C00387	HMDB00133
44	Hexanoylcarnitine	NA	HMDB00705
45	Hippurate	C01586	HMDB00714
46	Histidine	C00135	HMDB00177
47	Hypoxanthine	C00262	HMDB00157
48	IMP	C00130	HMDB00175
49	Inosine	C00294	HMDB00195
50	Isobutyryl-L-carnitine	NA	HMDB00736
51	Isoleucine	C00407	HMDB00172
52	Isovalerylcarnitine	NA	HMDB00688
53	Kynurenine	C01718	HMDB00684
54	Leucine	C00123	HMDB00687
55	Lysine	C00047	HMDB00182
56	Methionine	C00073	HMDB00696
57	MTA	NA	HMDB01173
58	Myoinositol	C00137	HMDB00211
59	N5-Formyl-THF	C03479	HMDB01562
60	NAD	C00003	HMDB00902
61	Neopterin	C05926	HMDB00845
62	Niacinamide	C00153	HMDB01406
63	Normetanephrine	C05589	HMDB00819
64	Octanoylcarnitine	C02838	HMDB00791
65	Ornithine	C00077	HMDB00214
66	Orotate	C00295	HMDB00226
67	Palmitoylcarnitine	C02990	HMDB00222
68	Pantothenate	C00864	HMDB00210
69	Phenylalanine	C00079	HMDB00159
70	Phosphoethanolamine	C00346	HMDB00224
71	Proline	C00148	HMDB00162
72	Propionylcarnitine	C03017	HMDB00824
73	SAH	C00021	HMDB00939
74	Serine	C00065	HMDB00187
75	Sorbitol	C00794	HMDB00247
76	Spermidine	C00315	HMDB01257
77	Stearoylcarnitine	NA	HMDB00848
78	Succinate	C00042	HMDB00254
79	Sucrose	C00089	HMDB00258
80	Taurine	C00245	HMDB00251

81	Taurochenodesoxycholate	C05465	HMDB00951
82	Taurocholate	C05122	HMDB00036
83	Tetradecanoylcarnitine	NA	HMDB05066
84	THF	C00101	HMDB01846
85	Threonine	C00188	HMDB00167
86	Trimethylamine N-oxide	C01104	HMDB00925
87	Tryptophan	C00078	HMDB00929
88	Tyrosine	C00082	HMDB00158
89	UDP Glucose	C00029	HMDB00286
90	Uracil	C00106	HMDB00300
91	Valine	C00183	HMDB00883
92	Xanthine	C00385	HMDB00292

Table A1: A list of all metabolites measured in this study and their KEGG (Kyoto Encyclopedia of Genes and Genoms) and HMDB (Human Metabolome Database) compound numbers.

Amino acids and derivatives	alanine, arginine, asparagine, aspartic acid, citrulline, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, kynurenine, leucine, lysine, methionine, ornithine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine
Acyl carnitines	acetylcarnitine, arachidyl carnitine, carnitine, decanoylcarnitine, dodecanoylcarnitine, hexanoylcarnitine, isobutyryl-l-carnitine, isovalerylcarnitine, octanoylcarnitine, palmitoylcarnitine, propionylcarnitine, stearoylcarnitine, tetradecanoylcarnitine
Folate cycle	5-methyl-THF, 5,10-metheny-THF, folic acid, N5-formyl-THF, THF
Nucleobases and related	adenine, adenosine, allantoin, AMP, cytidine, cytosine, deoxycytidine, deoxyuridine, guanosine, hypoxanthine, IMP, inosine, MTA, NAD, neopterin, SAH, uracil, xanthine
B-vitamins and degradation products	4-pyridoxate, niacinamide, orotate, pantothenate
Bile acids	chenodeoxycholate, choline, glycocholate, taurochenodesoxycholate, taurocholate
Sugars and related	D-ribose-5-P, glyceraldehyde, myoinositol, sorbitol, sucrose, UDP glucose
Carboxylic acids and amines	1-methylhistamine, 2-aminoisobutyrate, 3-hydroxanthranilate, acetoacetate, betaine, creatine, creatinine, dimethylglycine, guanidinoacetate, hippuric acid, spermidine, succinate
Other	G-glutamylcysteine, glutathione, normetanephrine, phosphoethanolamine, taurine, trimethylamine-N-oxide

Table A2: All analysed metabolites organised by chemical or biochemical compound classes.

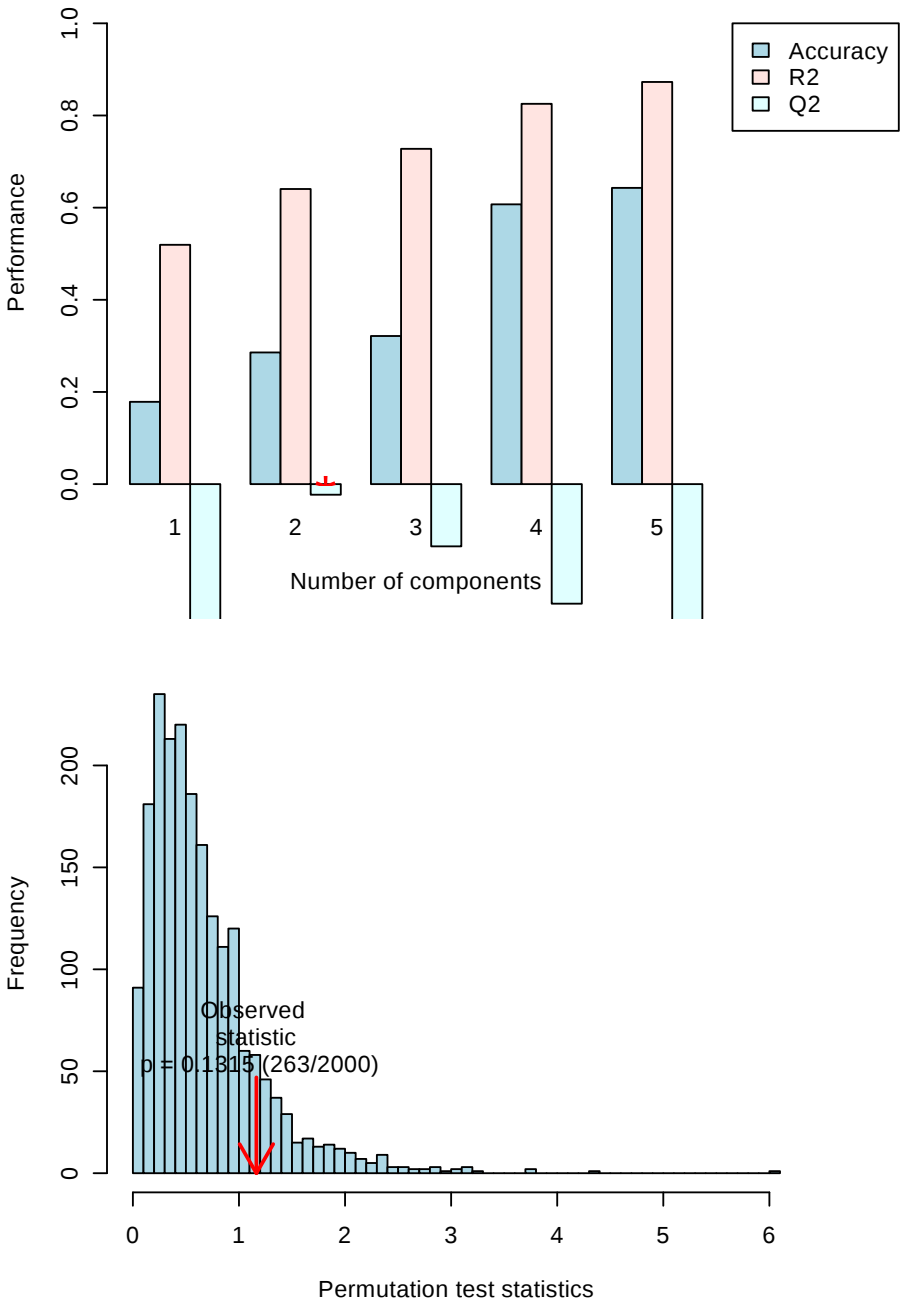


Figure A1: PLS-DA cross validation and permutations.

Metabolite	p.value	FDR	Fisher's LSD
Succinate	0.0005	0.027	Del - FGF21 KO; FGF21 KO Del - FGF21 KO; WT - FGF21 KO
Glutamate	0.0009	0.027	Del - FGF21 KO; Del - FGF21 KO Del; Del - WT
Betaine	0.0014	0.027	Del - FGF21 KO; Del - WT; FGF21 KO Del - FGF21 KO
Phosphoethanolamine	0.0015	0.027	Del - FGF21 KO; Del - FGF21 KO Del; Del - WT
Deoxyuridine	0.0018	0.027	FGF21 KO Del - Del; Del - WT; FGF21 KO - WT; FGF21 KO Del - WT
Creatine	0.0019	0.027	Del - FGF21 KO; FGF21 KO Del - FGF21 KO; WT - FGF21 KO
Guanidoacetate	0.0021	0.027	Del - FGF21 KO; FGF21 KO Del - FGF21 KO; WT - FGF21 KO
Threonine	0.0056	0.060	Del - FGF21 KO; Del - FGF21 KO Del; Del - WT
Palmitoylcarnitine	0.0065	0.060	Del - FGF21 KO; FGF21 KO Del - FGF21 KO; FGF21 KO Del - WT
Normetanephrine	0.0070	0.060	Del - FGF21 KO; FGF21 KO Del - FGF21 KO
Glutamine	0.0074	0.060	Del - FGF21 KO; FGF21 KO Del - FGF21 KO
Taurine	0.0078	0.060	Del - FGF21 KO; FGF21 KO Del - FGF21 KO; WT - FGF21 KO
Niacinamide	0.0088	0.062	Del - FGF21 KO; Del - FGF21 KO Del; Del - WT
Taurocholate	0.0096	0.063	Del - FGF21 KO; FGF21 KO Del - FGF21 KO

Table A3: The top 15 metabolites with the lowest p-value in ANOVA.

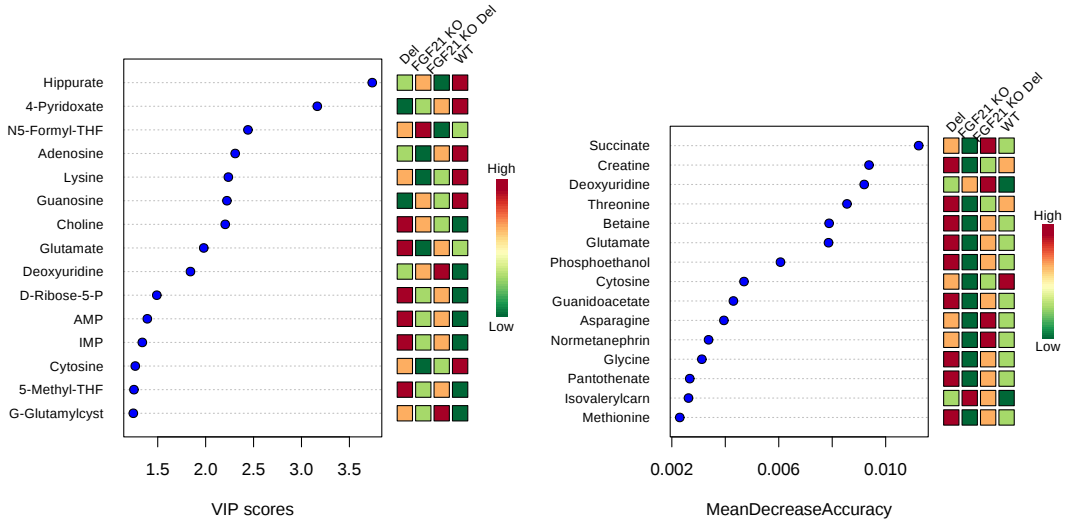


Figure A2: Metabolites with the most weight in classification in the PLS-DA and RF models ranked by (left) variable importance in projection (VIP) scores in PLS-DA and (right) mean decrease accuracy (MDA) in RF.

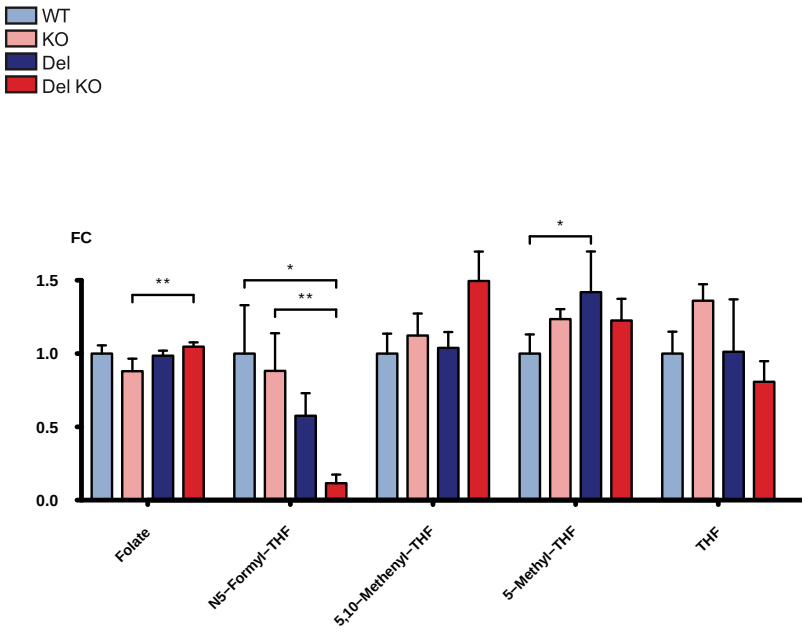


Figure A3: Folate cycle intermediates. Data presented as fold change compared to WT. THF = tetrahydrofolate. The significances between each pair of groups were determined by Student's t-test. * = $p < 0.05$, ** = $p < 0.01$

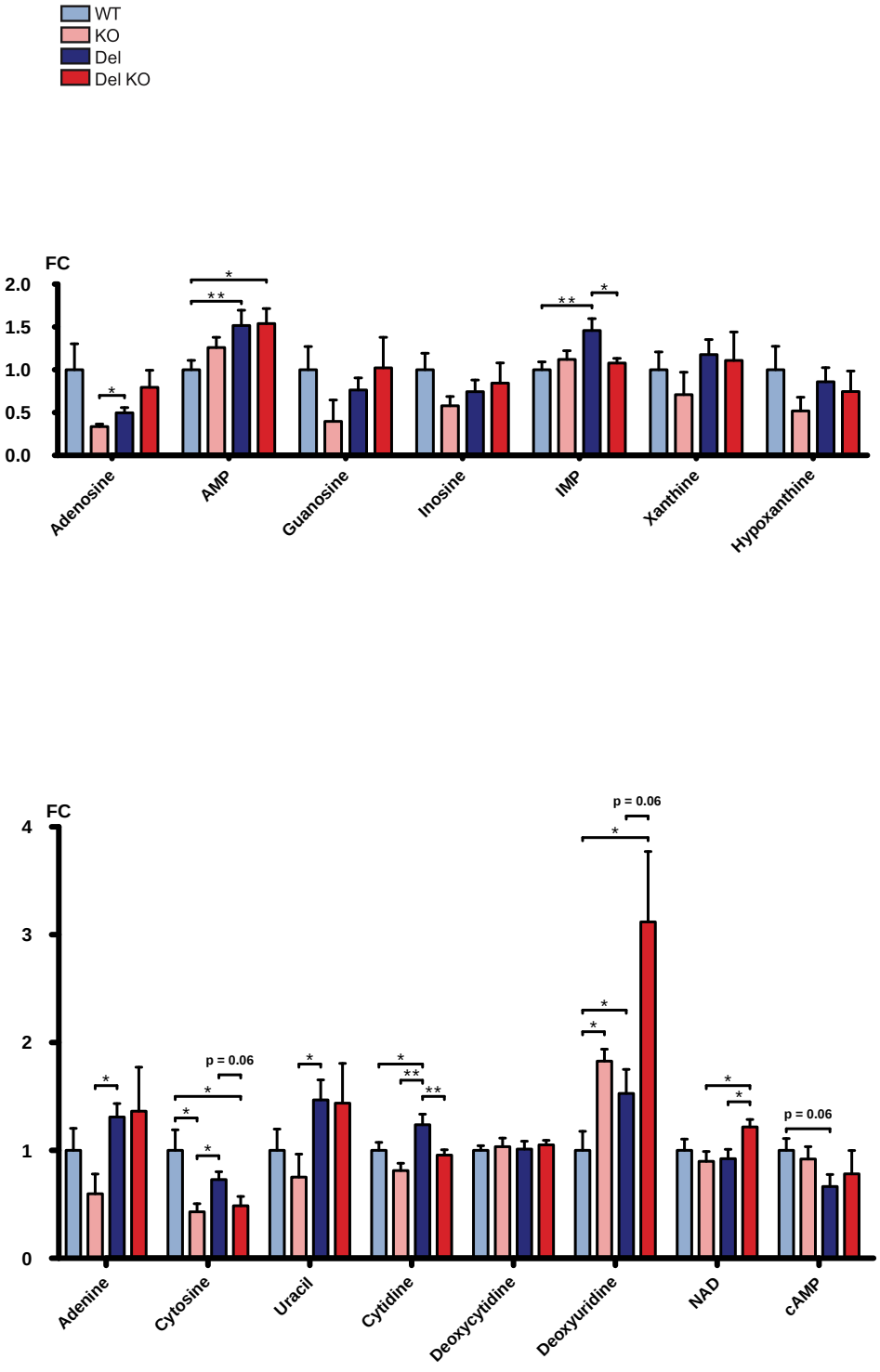
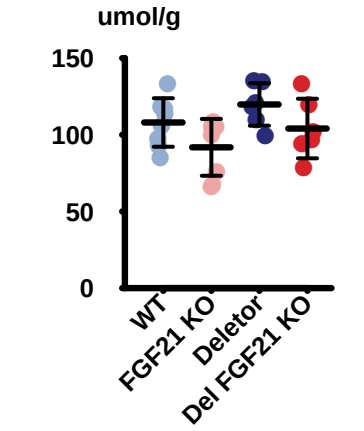


Figure A4: Metabolites related to nucleotides. AMP = adenosine monophosphate, IMP = inosine monophosphate. The significances between each pair of groups were determined by Student's t-test. * = $p < 0.05$, ** = $p < 0.01$

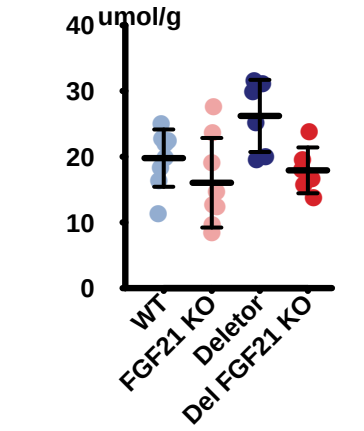
Acetoacetate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1659	0.1413	0.0072	0.6932	0.0814	0.2594

fc_wt	fc_KO	fc_del	fc_delKO
1	0.85	1.109	0.963

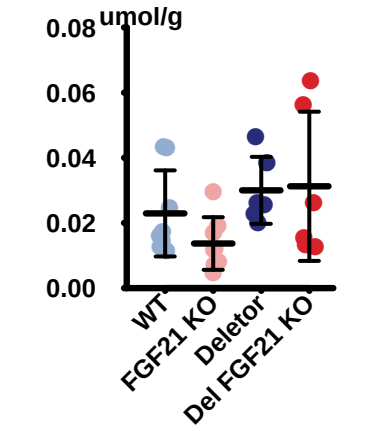
Acetylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0416	0.0133	0.0095	0.3923	0.2147	0.5144

fc_wt	fc_KO	fc_del	fc_delKO
1	0.81	1.325	0.906

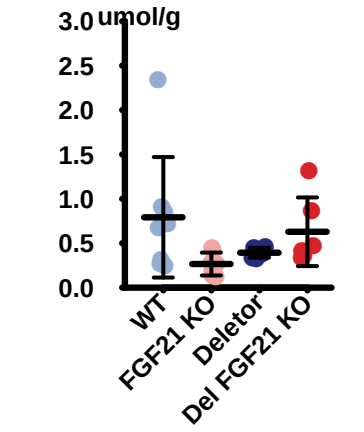
Adenine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2815	0.9069	0.0102	0.4503	0.1183	0.1231

fc_wt	fc_KO	fc_del	fc_delKO
1	0.597	1.309	1.363

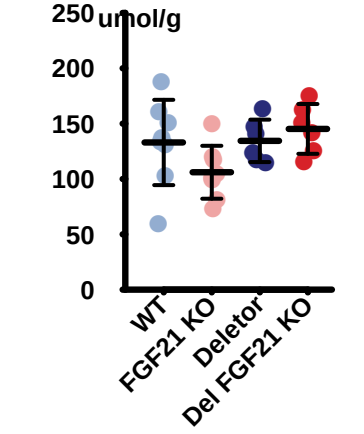
Adenosine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1411	0.1967	0.0443	0.5817	0.0658	0.0706

fc_wt	fc_KO	fc_del	fc_delKO
1	0.336	0.496	0.794

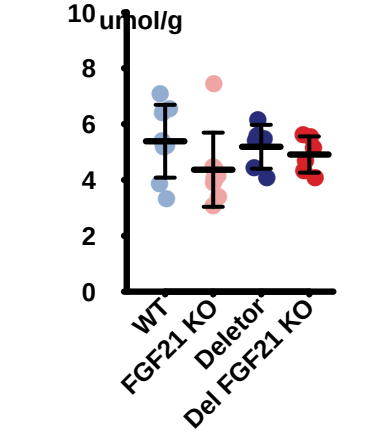
Alanine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.9278	0.3928	0.0301	0.471	0.1197	0.0092

fc_wt	fc_KO	fc_del	fc_delKO
1	0.798	1.011	1.092

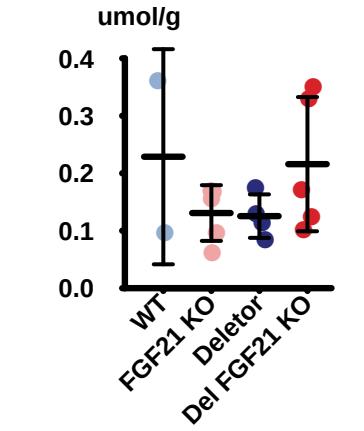
Allantoin



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.7312	0.52	0.1752	0.3909	0.1438	0.3356

fc_wt	fc_KO	fc_del	fc_delKO
1	0.811	0.963	0.912

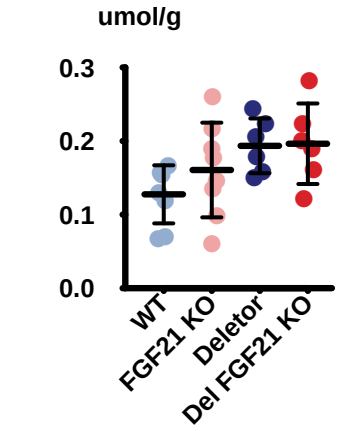
Aminoadipate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.577	0.1645	0.857	0.9394	0.5933	0.1893

fc_wt	fc_KO	fc_del	fc_delKO
1	0.572	0.548	0.943

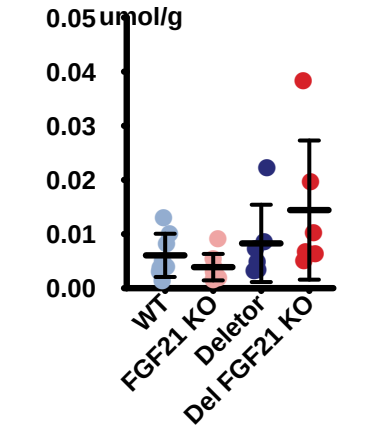
AMP



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0083	0.9182	0.2543	0.029	0.2398	0.2855

fc_wt	fc_KO	fc_del	fc_delKO
1	1.259	1.517	1.411

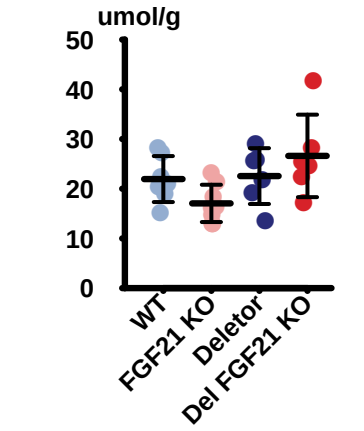
Arachidyl carnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5157	0.336	0.1998	0.177	0.2134	0.1013

fc_wt	fc_KO	fc_del	fc_delKO
1	0.641	1.365	2.38

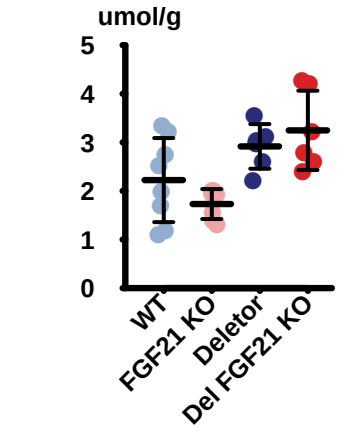
Arginine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.8387	0.3475	0.0707	0.258	0.0464	0.0363

fc_wt	fc_KO	fc_del	fc_delKO
1	0.777	1.027	1.213

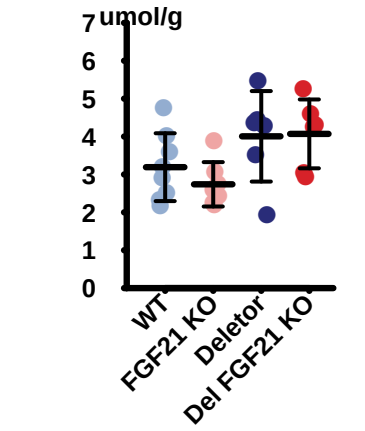
Asparagine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0793	0.4144	5e−04	0.0445	0.1666	0.0047

fc_wt	fc_KO	fc_del	fc_delKO
1	0.778	1.312	1.46

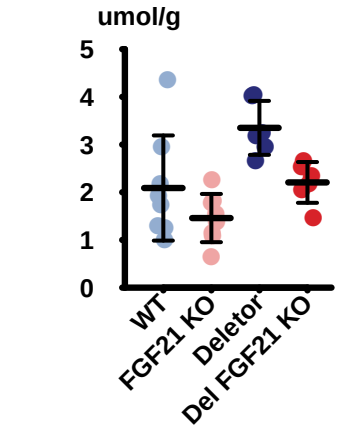
Aspartate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1949	0.9187	0.0499	0.0995	0.2645	0.0144

fc_wt	fc_KO	fc_del	fc_delKO
1	0.858	1.255	1.275

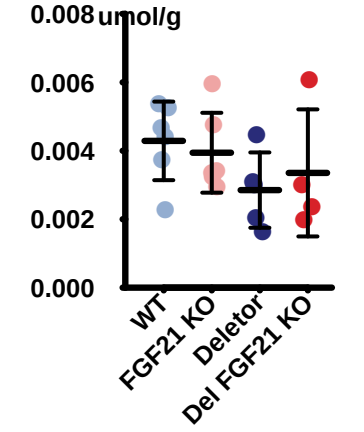
Betaine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0179	0.0031	1e−04	0.7894	0.173	0.0115

fc_wt	fc_KO	fc_del	fc_delKO
1	0.698	1.604	1.056

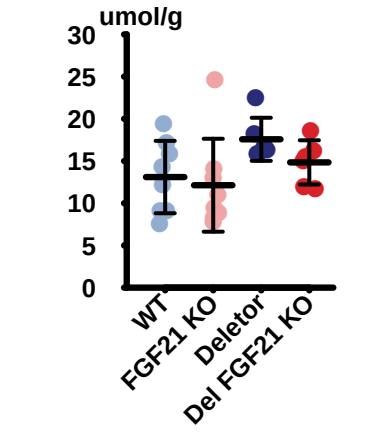
cAMP



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0642	0.6543	0.146	0.4132	0.6159	0.5982

fc_wt	fc_KO	fc_del	fc_delKO
1	0.919	0.665	0.782

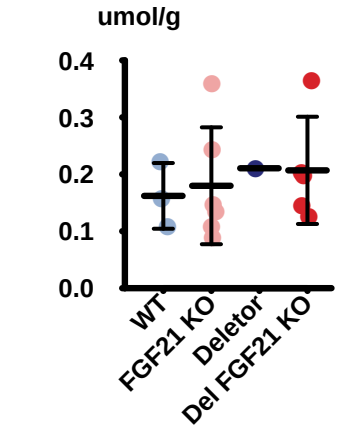
Carnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0323	0.0967	0.0326	0.3679	0.7011	0.2498

fc_wt	fc_KO	fc_del	fc_delKO
1	0.926	1.342	1.133

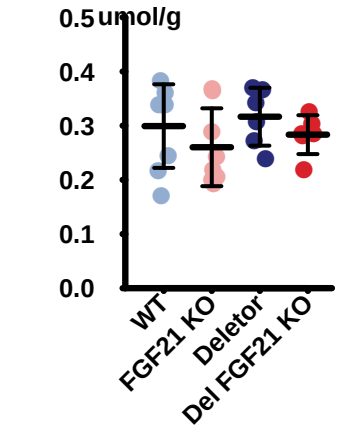
Carnosine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO

fc_wt	fc_KO	fc_del	fc_delKO
1	1.109	1.3	1.277

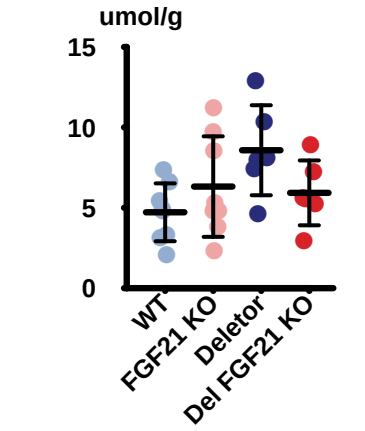
Chenodeoxycholate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.6291	0.2402	0.1183	0.6207	0.3129	0.4452

fc_wt	fc_KO	fc_del	fc_delKO
1	0.87	1.058	1.207

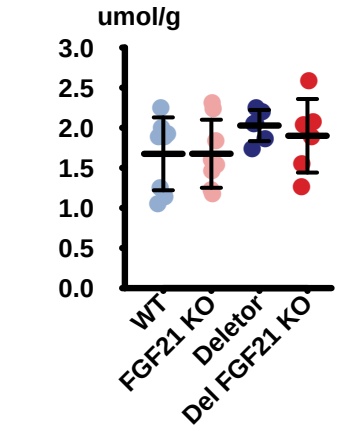
Choline



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0183	0.0921	0.1818	0.2721	0.2352	0.781

fc_wt	fc_KO	fc_del	fc_delKO
1	1.339	1.818	1.256

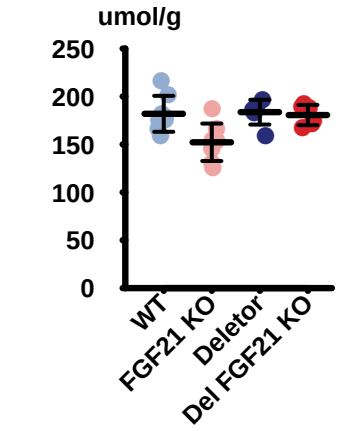
Citrulline



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.077	0.5497	0.064	0.3813	0.9959	0.3717

fc_wt	fc_KO	fc_del	fc_delKO
1	1.001	1.211	1.134

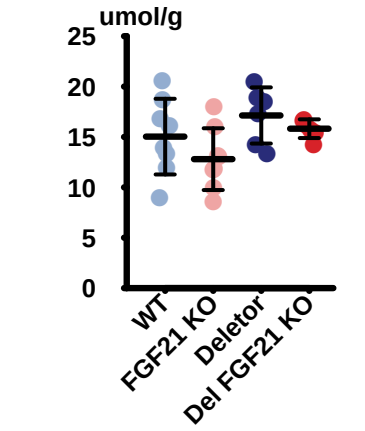
Creatine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.8421	0.6663	0.0036	0.8713	0.0078	0.005

fc_wt	fc_KO	fc_del	fc_delKO
1	0.837	1.009	0.993

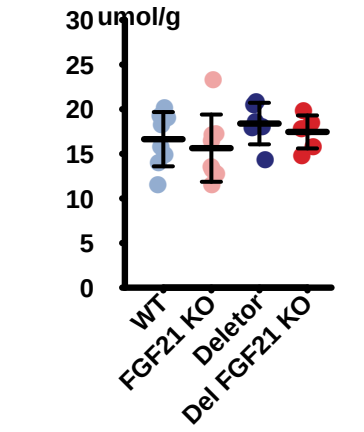
Creatinine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.253	0.3149	0.018	0.5843	0.2139	0.0282

fc_wt	fc_KO	fc_del	fc_delKO
1	0.851	1.14	1.052

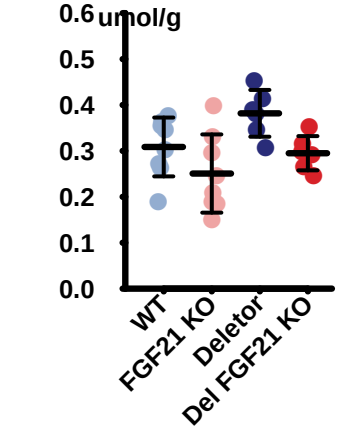
Cysteine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2458	0.4584	0.1187	0.5475	0.5673	0.2608

fc_wt	fc_KO	fc_del	fc_delKO
1	0.94	1.105	1.049

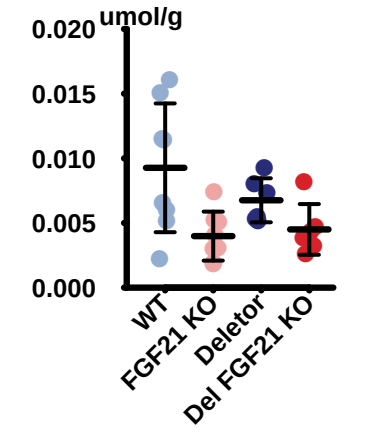
Cytidine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0348	0.008	0.004	0.6264	0.1485	0.2189

fc_wt	fc_KO	fc_del	fc_delKO
1	0.812	1.237	0.956

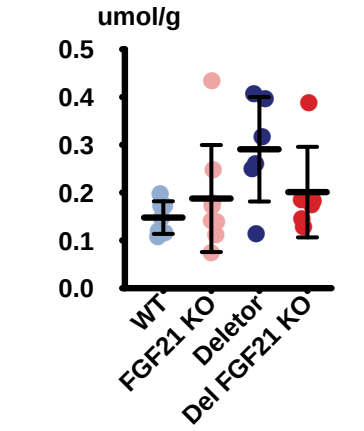
Cytosine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.217	0.0599	0.0145	0.0343	0.0206	0.6291

fc_wt	fc_KO	fc_del	fc_delKO
1	0.43	0.729	0.486

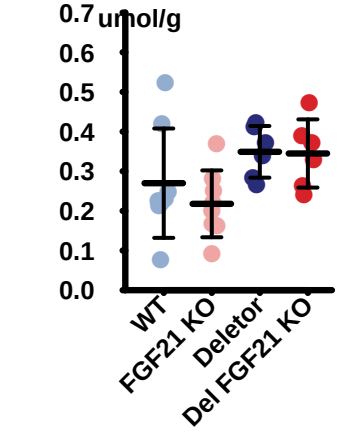
D−Ribose−5−P



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0225	0.1608	0.1118	0.2381	0.3651	0.8121

fc_wt	fc_KO	fc_del	fc_delKO
1	1.27	1.968	1.361

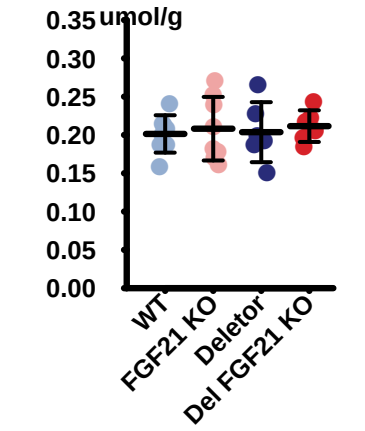
Decanoylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1837	0.9267	0.0066	0.2369	0.3805	0.0189

fc_wt	fc_KO	fc_del	fc_delKO
1	0.807	1.293	1.28

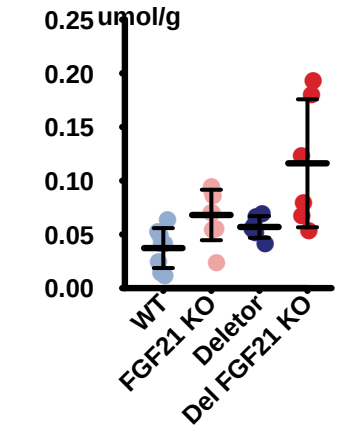
Deoxycytidine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.902	0.6752	0.8384	0.4154	0.6953	0.847

fc_wt	fc_KO	fc_del	fc_delKO
1	1.034	1.011	1.051

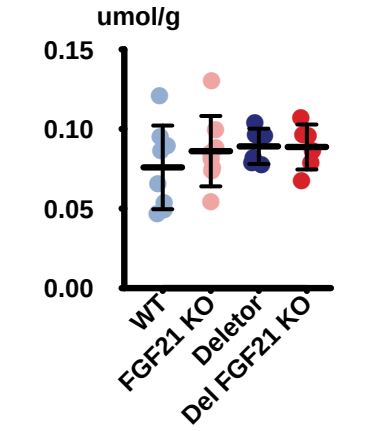
Deoxyuridine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0274	0.0587	0.2561	0.0215	0.0119	0.109

fc_wt	fc_KO	fc_del	fc_delKO
1	1.827	1.528	3.118

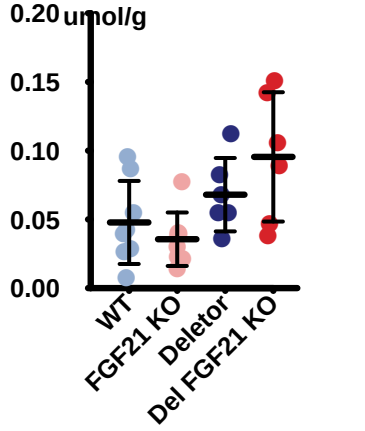
Dimethylglycine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2305	0.9579	0.744	0.266	0.4163	0.7912

fc_wt	fc_KO	fc_del	fc_delKO
1	1.134	1.174	1.169

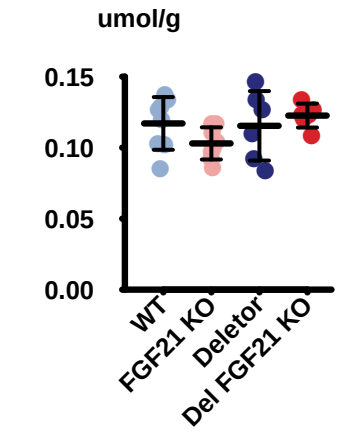
Dodecanoylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2112	0.2485	0.0337	0.0616	0.3561	0.0247

fc_wt	fc_KO	fc_del	fc_delKO
1	0.745	1.423	1.999

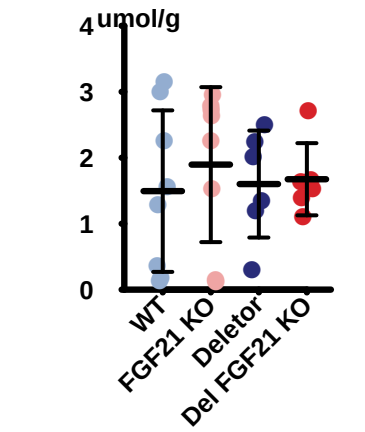
Folate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.8906	0.52	0.2895	0.4732	0.0929	0.003

fc_wt	fc_KO	fc_del	fc_delKO
1	0.88	0.986	1.047

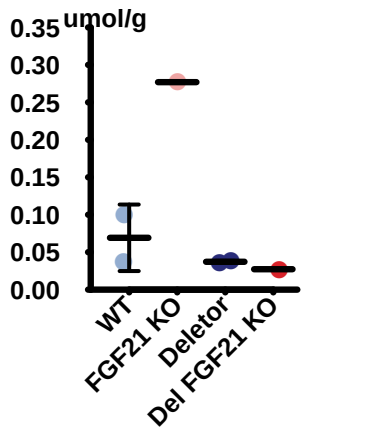
G−Glutamylcysteine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.8474	0.859	0.5891	0.7189	0.5138	0.648

fc_wt	fc_KO	fc_del	fc_delKO
1	1.269	1.072	1.121

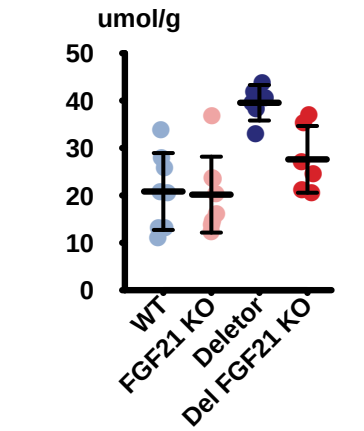
Glucuronate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO

fc_wt	fc_KO	fc_del	fc_delKO
1	4.002	0.537	0.394

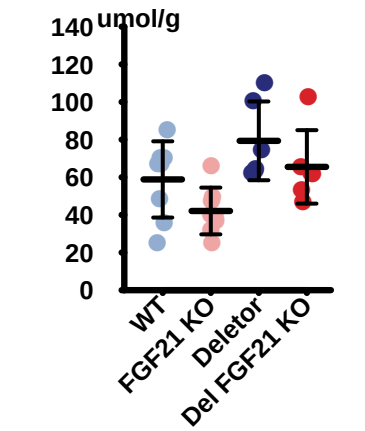
Glutamate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
2e−04	0.0069	1e−04	0.1215	0.8721	0.0908

fc_wt	fc_KO	fc_del	fc_delKO
1	0.968	1.899	1.326

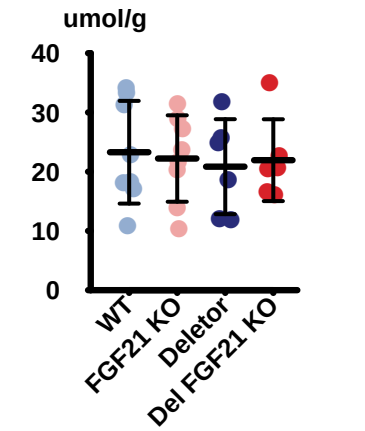
Glutamine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0938	0.2639	0.0052	0.5465	0.07	0.033

fc_wt	fc_KO	fc_del	fc_delKO
1	0.715	1.348	1.113

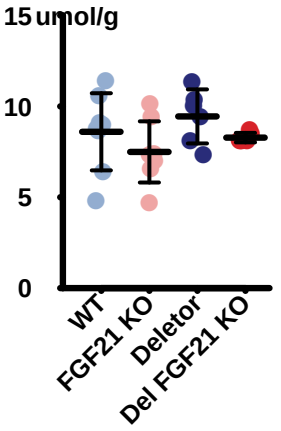
Glutathione



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5972	0.8048	0.7478	0.753	0.795	0.9429

fc_wt	fc_KO	fc_del	fc_delKO
1	0.954	0.895	0.942

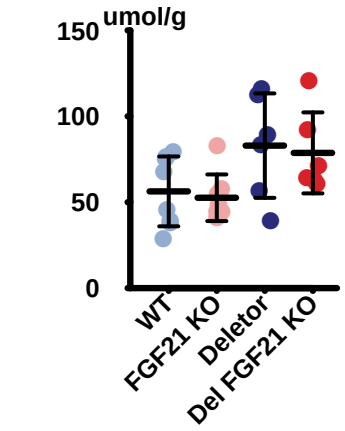
Glyceraldehyde



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.398	0.1136	0.0411	0.6846	0.268	0.232

fc_wt	fc_KO	fc_del	fc_delKO
1	0.871	1.098	0.963

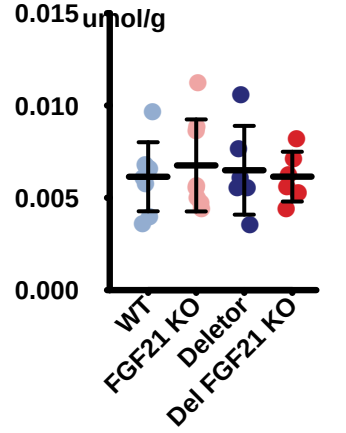
Glycine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0993	0.7935	0.0595	0.0919	0.6755	0.0434

fc_wt	fc_KO	fc_del	fc_delKO
1	0.934	1.473	1.398

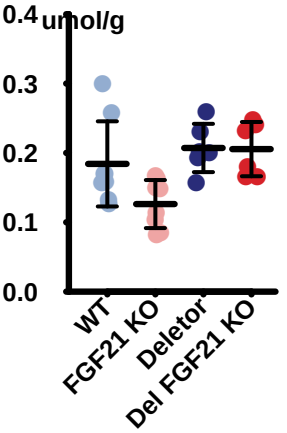
Glycocholate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.7728	0.767	0.8461	0.9939	0.5869	0.5706

fc_wt	fc_KO	fc_del	fc_delKO
1	1.1	1.057	1.001

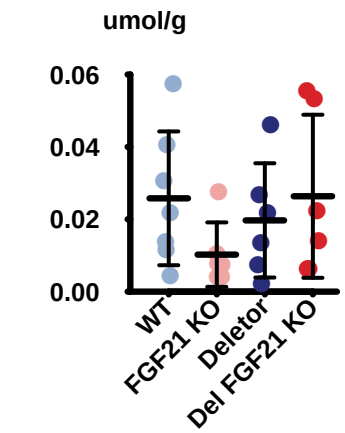
Guanidoacetate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.3946	0.9405	0.0013	0.4446	0.0401	0.0027

fc_wt	fc_KO	fc_del	fc_delKO
1	0.686	1.125	1.116

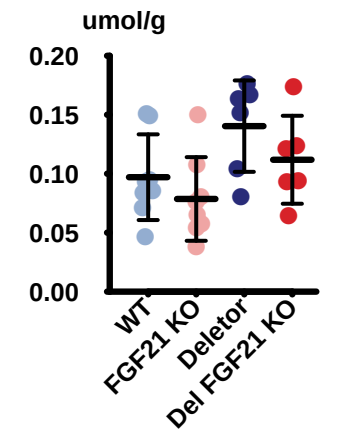
Guanosine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5343	0.5676	0.2374	0.9618	0.08	0.1502

fc_wt	fc_KO	fc_del	fc_delKO
1	0.397	0.764	1.022

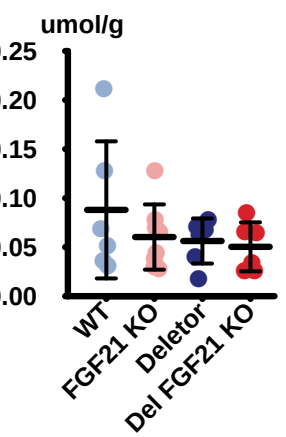
Hexanoylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0582	0.2234	0.0116	0.4735	0.3216	0.1211

fc_wt	fc_KO	fc_del	fc_delKO
1	0.81	1.446	1.152

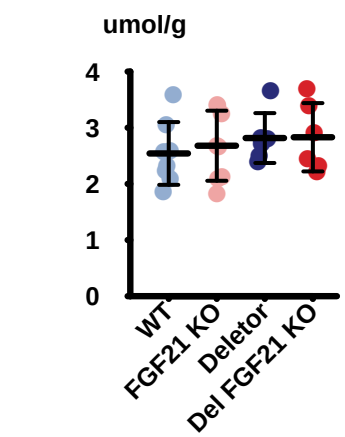
Hippurate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.3314	0.6764	0.7936	0.2585	0.4009	0.5332

fc_wt	fc_KO	fc_del	fc_delKO
1	0.685	0.64	0.572

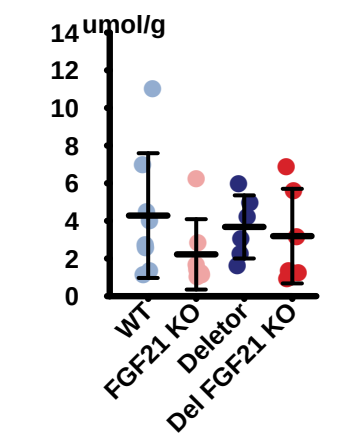
Histidine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.3265	0.9642	0.6405	0.3839	0.6496	0.6582

fc_wt	fc_KO	fc_del	fc_delKO
1	1.054	1.108	1.11

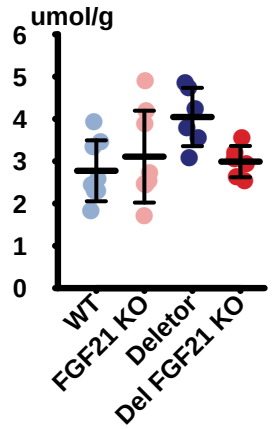
Hypoxanthine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.6658	0.7017	0.1666	0.497	0.1596	0.4561

fc_wt	fc_KO	fc_del	fc_delKO
1	0.519	0.859	0.745

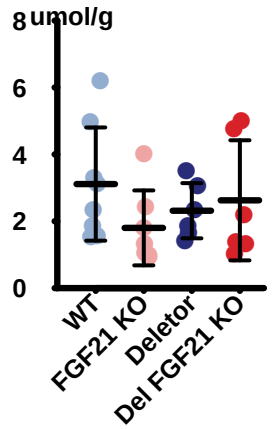
IMP



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0063	0.0115	0.073	0.4794	0.4818	0.7822

fc_wt	fc_KO	fc_del	fc_delKO
1	1.121	1.458	1.078

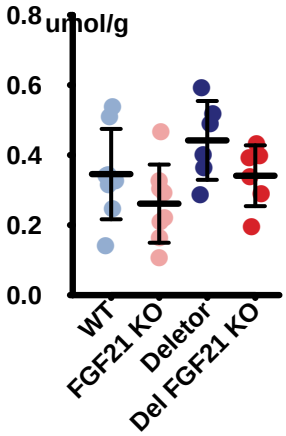
Inosine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2709	0.7111	0.3641	0.618	0.0988	0.3585

fc_wt	fc_KO	fc_del	fc_delKO
1	0.579	0.744	0.844

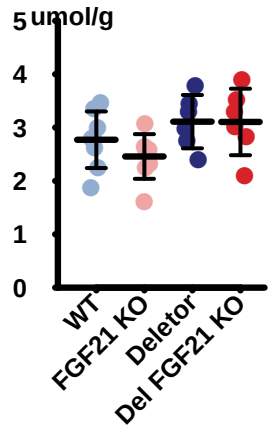
Isobutyryl−L−carnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1638	0.1147	0.0126	0.9369	0.184	0.159

fc_wt	fc_KO	fc_del	fc_delKO
1	0.756	1.278	0.986

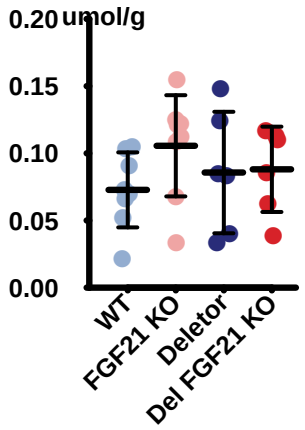
Isoleucine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2463	0.9879	0.0276	0.3161	0.2116	0.0579

fc_wt	fc_KO	fc_del	fc_delKO
1	0.887	1.122	1.12

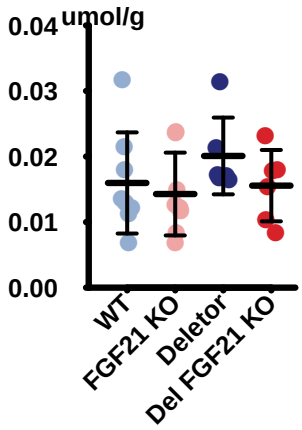
Isovalerylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5551	0.9179	0.403	0.3698	0.0695	0.3649

fc_wt	fc_KO	fc_del	fc_delKO
1	1.451	1.177	1.21

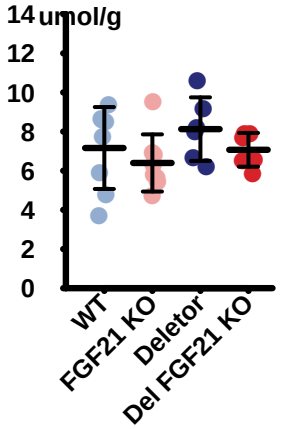
Kynurenine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2783	0.1954	0.1026	0.9089	0.6397	0.6925

fc_wt	fc_KO	fc_del	fc_delKO
1	0.894	1.258	0.974

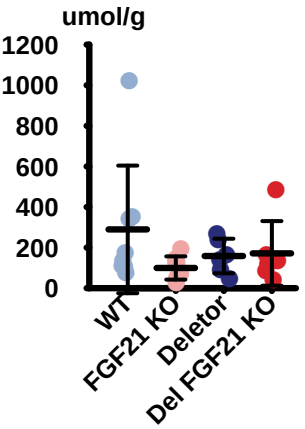
Leucine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.3507	0.1987	0.0661	0.9132	0.413	0.304

fc_wt	fc_KO	fc_del	fc_delKO
1	0.893	1.135	0.987

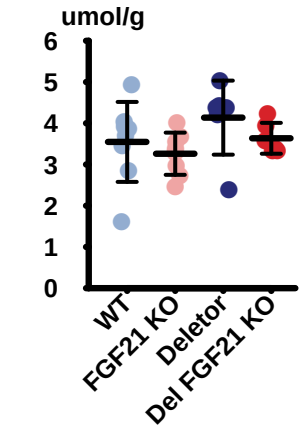
Lysine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2919	0.8649	0.1801	0.3789	0.1339	0.3313

fc_wt	fc_KO	fc_del	fc_delKO
1	0.344	0.547	0.551

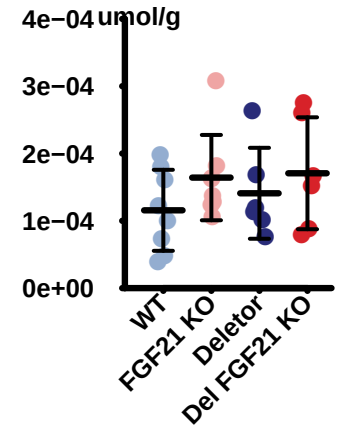
Methionine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2649	0.2494	0.0672	0.8204	0.4751	0.1398

fc_wt	fc_KO	fc_del	fc_delKO
1	0.919	1.166	1.025

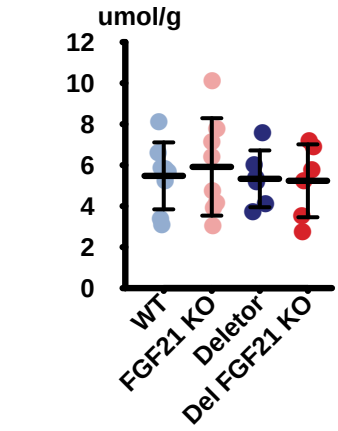
MTA



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.4841	0.5129	0.5264	0.2039	0.1383	0.8779

fc_wt	fc_KO	fc_del	fc_delKO
1	1.42	1.219	1.476

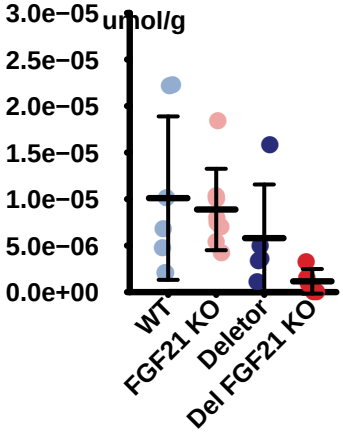
Myoinositol



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.8616	0.9186	0.577	0.8004	0.6752	0.5528

fc_wt	fc_KO	fc_del	fc_delKO
1	1.08	0.974	0.956

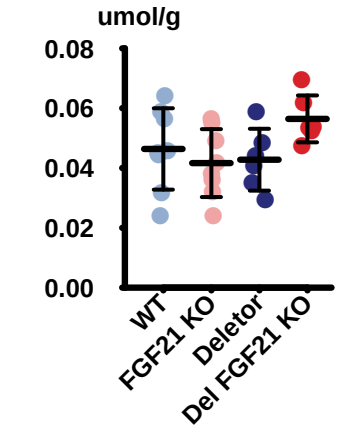
N5−Formyl−THF



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.3329	0.1488	0.3394	0.0364	0.7531	0.0013

fc_wt	fc_KO	fc_del	fc_delKO
1	0.882	0.576	0.115

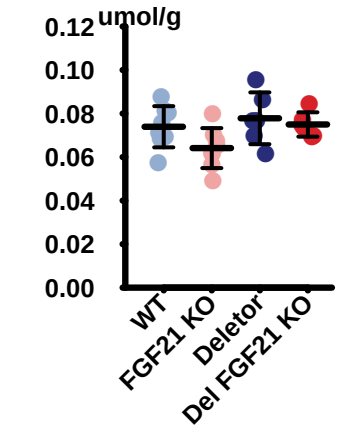
NAD



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5846	0.0291	0.8489	0.1086	0.4631	0.0139

fc_wt	fc_KO	fc_del	fc_delKO
1	0.898	0.922	1.217

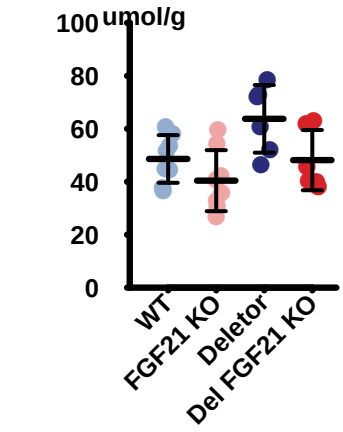
Neopterin



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5254	0.6123	0.0434	0.8007	0.054	0.0186

fc_wt	fc_KO	fc_del	fc_delKO
1	0.867	1.053	1.014

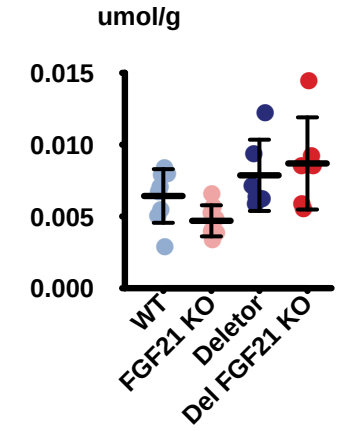
Niacinamide



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0363	0.05	0.0053	0.9391	0.1363	0.235

fc_wt	fc_KO	fc_del	fc_delKO
1	0.831	1.312	0.991

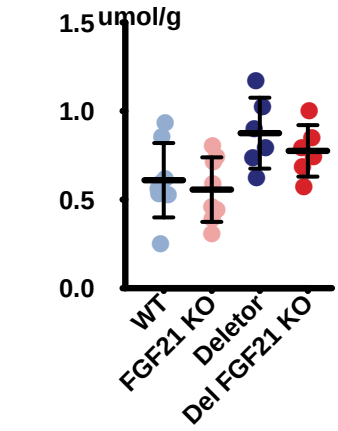
Normetanephrine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2644	0.6292	0.0243	0.1635	0.0442	0.0271

fc_wt	fc_KO	fc_del	fc_delKO
1	0.731	1.224	1.353

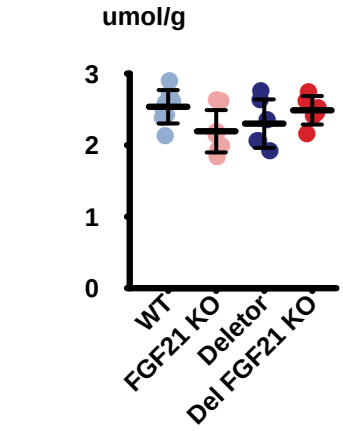
Octanoylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0344	0.3476	0.0116	0.107	0.5985	0.0283

fc_wt	fc_KO	fc_del	fc_delKO
1	0.913	1.436	1.37

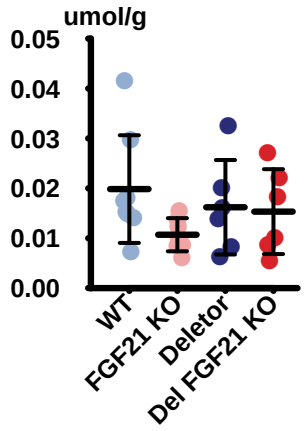
Ornithine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1796	0.2781	0.5517	0.6791	0.023	0.0472

fc_wt	fc_KO	fc_del	fc_delKO
1	0.865	0.907	0.981

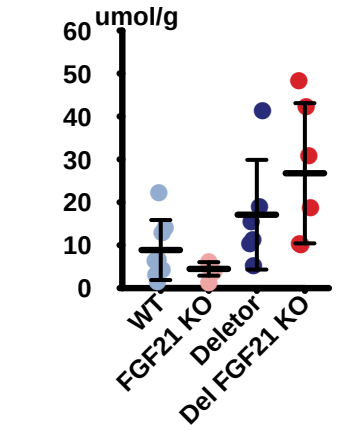
Orotate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5155	0.8697	0.2238	0.3985	0.0505	0.2545

fc_wt	fc_KO	fc_del	fc_delKO
1	0.539	0.816	0.772

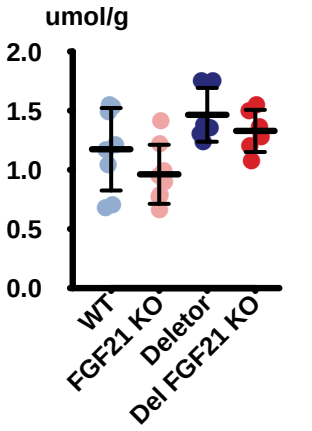
Palmitoylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1957	0.282	0.0599	0.0433	0.1227	0.0203

fc_wt	fc_KO	fc_del	fc_delKO
1	0.504	1.927	3.016

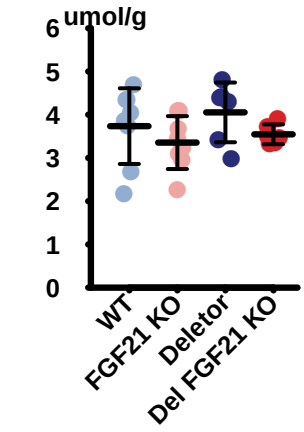
Pantothenate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.084	0.2771	0.0022	0.3023	0.1864	0.0075

fc_wt	fc_KO	fc_del	fc_delKO
1	0.82	1.248	1.132

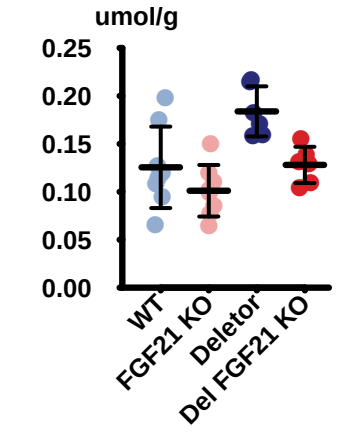
Phenylalanine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.4621	0.1378	0.0774	0.5736	0.3328	0.4377

fc_wt	fc_KO	fc_del	fc_delKO
1	0.898	1.085	0.949

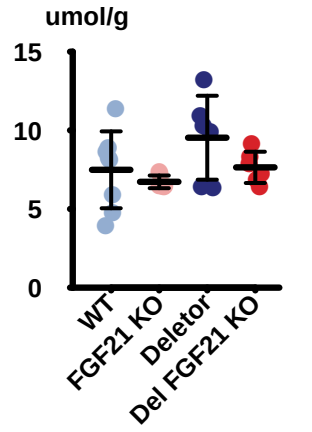
Phosphoethanolamine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0084	0.0021	1e−04	0.8853	0.1945	0.0481

fc_wt	fc_KO	fc_del	fc_delKO
1	0.805	1.465	1.02

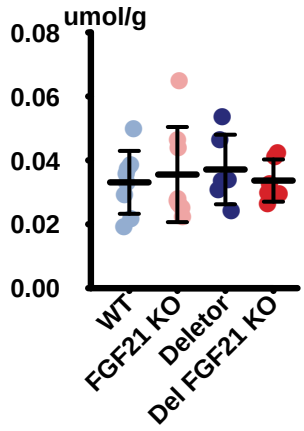
Proline



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1718	0.1532	0.0495	0.8734	0.4111	0.0752

fc_wt	fc_KO	fc_del	fc_delKO
1	0.898	1.272	1.021

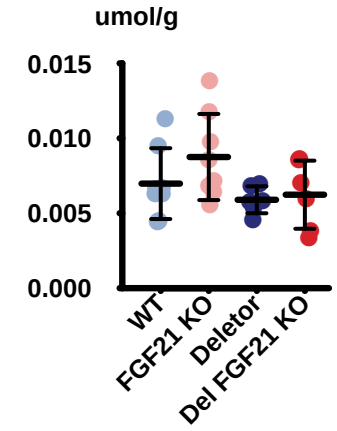
Propionylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.4919	0.523	0.822	0.9018	0.7057	0.7565

fc_wt	fc_KO	fc_del	fc_delKO
1	1.074	1.121	1.017

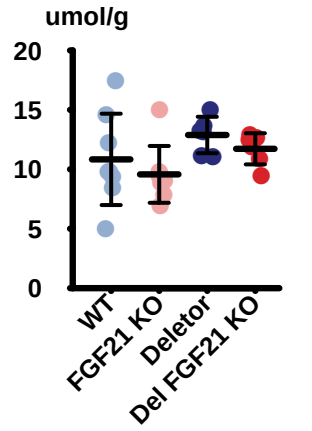
SAH



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2655	0.7441	0.0275	0.5651	0.1995	0.0927

fc_wt	fc_KO	fc_del	fc_delKO
1	1.254	0.845	0.994

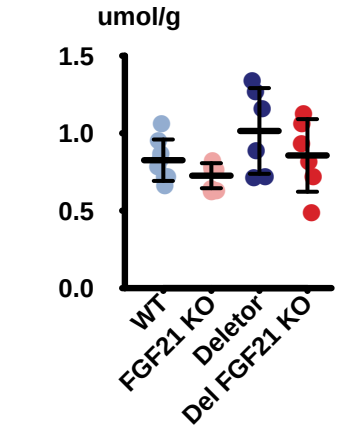
Serine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2012	0.1887	0.0085	0.5581	0.4461	0.0543

fc_wt	fc_KO	fc_del	fc_delKO
1	0.884	1.189	1.082

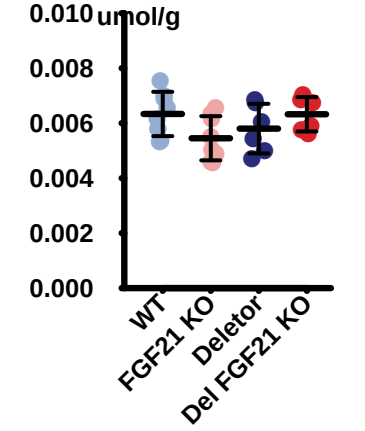
Sorbitol



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1685	0.3116	0.0504	0.7805	0.0963	0.2389

fc_wt	fc_KO	fc_del	fc_delKO
1	0.879	1.229	1.037

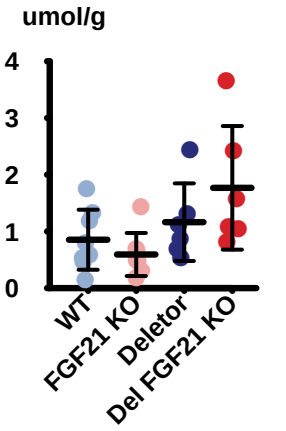
Spermidine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2783	0.2747	0.4688	0.9763	0.0454	0.0415

fc_wt	fc_KO	fc_del	fc_delKO
1	0.86	0.916	0.998

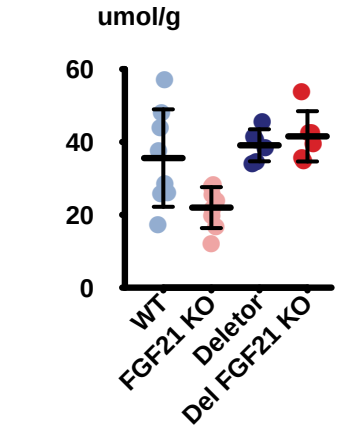
Stearoylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.3803	0.2811	0.1066	0.1012	0.2791	0.0453

fc_wt	fc_KO	fc_del	fc_delKO
1	0.695	1.363	2.073

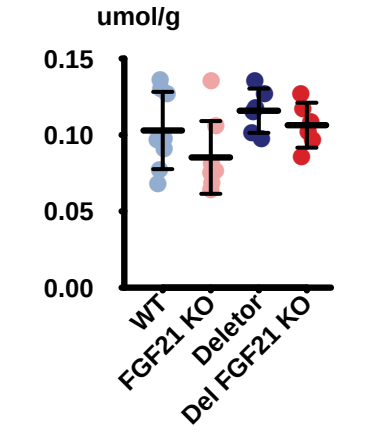
Succinate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5053	0.4815	0	0.3015	0.0255	2e−04

fc_wt	fc_KO	fc_del	fc_delKO
1	0.618	1.099	1.168

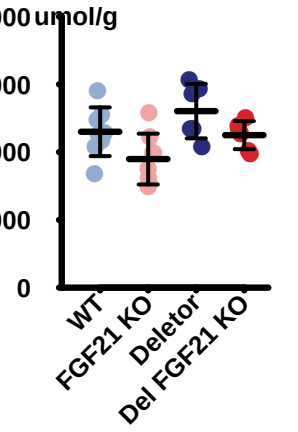
Sucrose



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2521	0.2879	0.012	0.7528	0.1729	0.0643

fc_wt	fc_KO	fc_del	fc_delKO
1	0.829	1.126	1.034

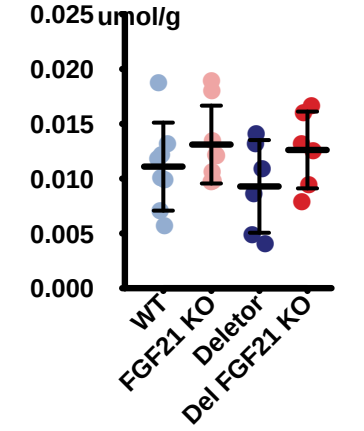
Taurine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1729	0.0939	0.0069	0.7504	0.0457	0.0457

fc_wt	fc_KO	fc_del	fc_delKO
1	0.825	1.132	0.978

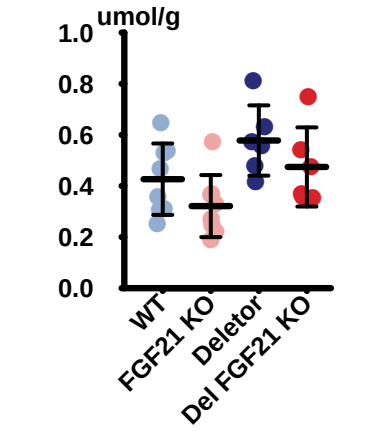
Taurochenodesoxycholat



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.4366	0.1698	0.1046	0.466	0.3064	0.7999

fc_wt	fc_KO	fc_del	fc_delKO
1	1.181	0.837	1.137

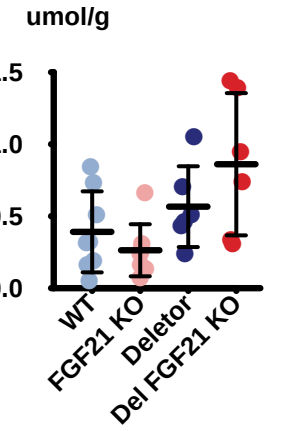
Taurocholate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0678	0.2491	0.0045	0.5634	0.1305	0.0751

fc_wt	fc_KO	fc_del	fc_delKO
1	0.753	1.355	1.512

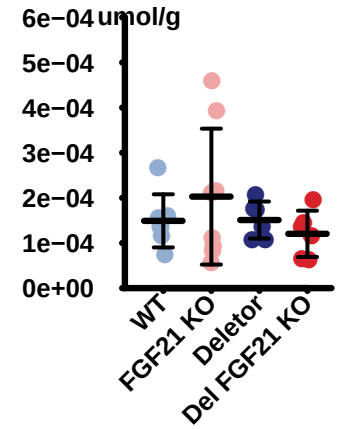
Tetradecanoylcarnitine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.2731	0.24	0.0496	0.0729	0.3013	0.0301

fc_wt	fc_KO	fc_del	fc_delKO
1	0.673	1.448	2.201

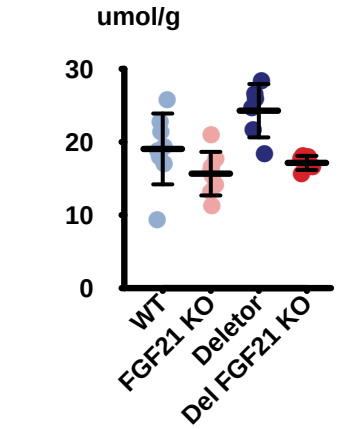
THF



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.9484	0.2827	0.3785	0.368	0.3751	0.1834

fc_wt	fc_KO	fc_del	fc_delKO
1	1.361	1.012	0.807

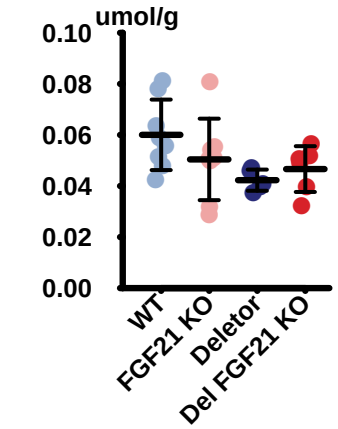
Threonine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.04	0.0041	9e−04	0.3105	0.1193	0.2227

fc_wt	fc_KO	fc_del	fc_delKO
1	0.822	1.275	0.9

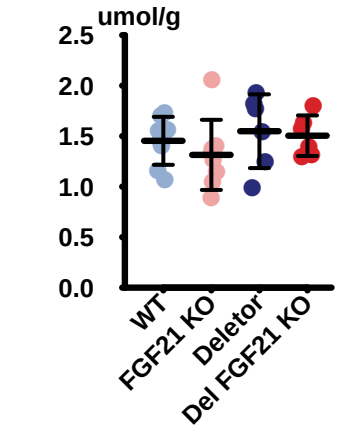
Trimethylamine N−oxide



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0081	0.3226	0.2061	0.0486	0.2185	0.5846

fc_wt	fc_KO	fc_del	fc_delKO
1	0.84	0.704	0.777

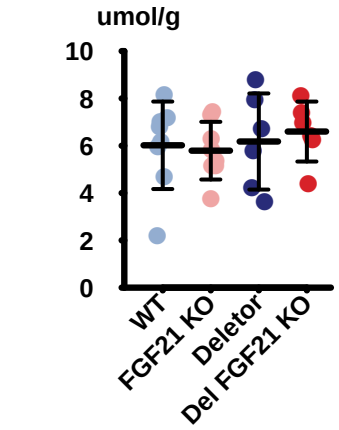
Tryptophan



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.5928	0.8002	0.2511	0.6737	0.3673	0.2238

fc_wt	fc_KO	fc_del	fc_delKO
1	0.904	1.066	1.035

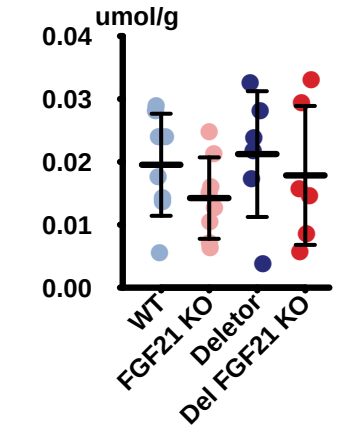
Tyrosine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.8826	0.6776	0.6899	0.4994	0.7762	0.2559

fc_wt	fc_KO	fc_del	fc_delKO
1	0.962	1.027	1.097

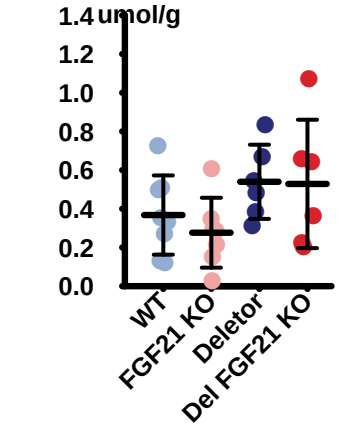
UDP Glucose



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.7405	0.5895	0.1728	0.7592	0.172	0.4972

fc_wt	fc_KO	fc_del	fc_delKO
1	0.729	1.087	0.913

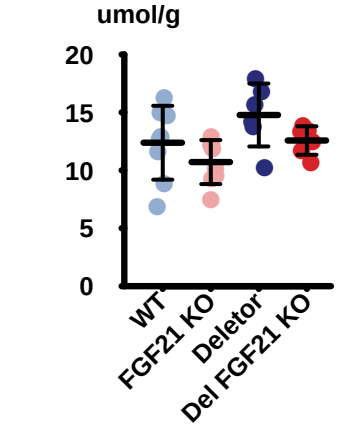
Uracil



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1344	0.9453	0.0288	0.3265	0.3757	0.1379

fc_wt	fc_KO	fc_del	fc_delKO
1	0.752	1.468	1.437

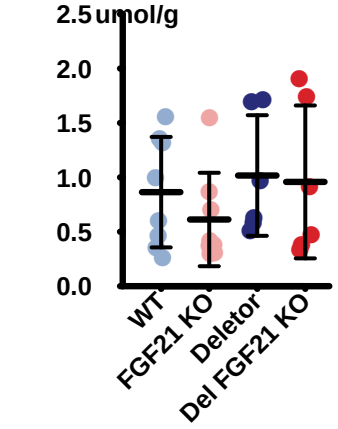
Valine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1566	0.1135	0.0128	0.8802	0.228	0.0465

fc_wt	fc_KO	fc_del	fc_delKO
1	0.865	1.193	1.331

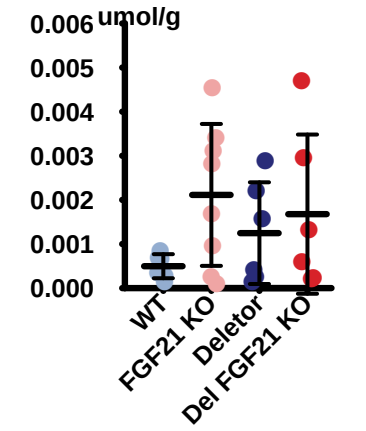
Xanthine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.6078	0.8754	0.1714	0.7877	0.3033	0.3193

fc_wt	fc_KO	fc_del	fc_delKO
1	0.709	1.177	1.109

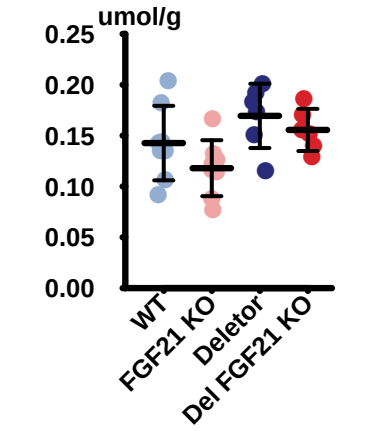
1−Methylhistamine



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.176	0.6348	0.2631	0.1715	0.025	0.6492

fc_wt	fc_KO	fc_del	fc_delKO
1	4.253	2.509	3.375

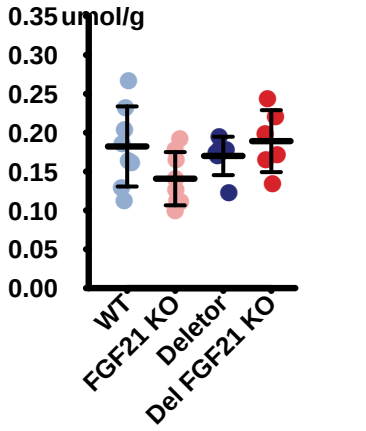
2−Aminoisobutyrate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1701	0.395	0.0098	0.4209	0.1511	0.0129

fc_wt	fc_KO	fc_del	fc_delKO
1	0.827	1.187	1.091

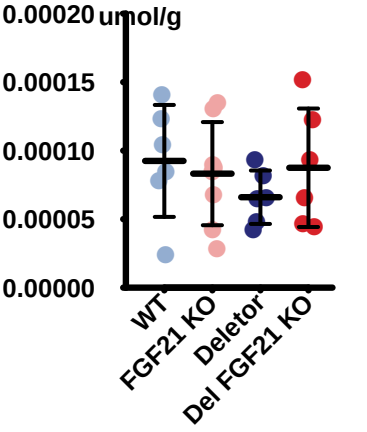
3−Hydroxanthranilate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.571	0.3484	0.0876	0.7843	0.0819	0.0387

fc_wt	fc_KO	fc_del	fc_delKO
1	0.772	0.933	1.038

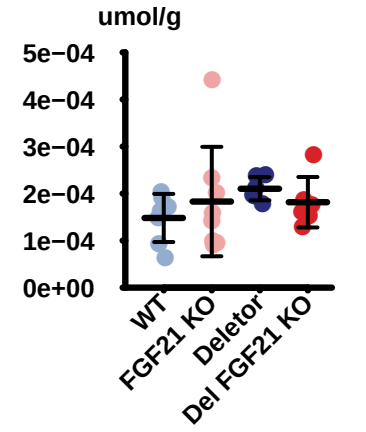
4−Pyridoxate



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.1934	0.304	0.2915	0.8397	0.6706	0.8497

fc_wt	fc_KO	fc_del	fc_delKO
1	0.899	0.713	0.945

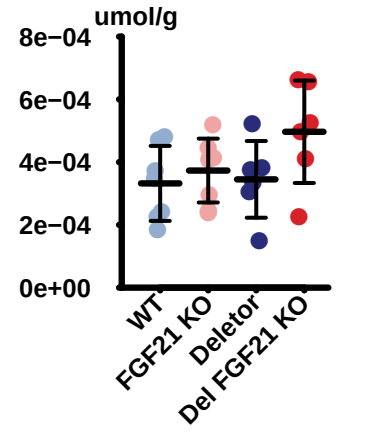
5−Methyl−THF



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.0194	0.2737	0.5387	0.2779	0.4611	0.9765

fc_wt	fc_KO	fc_del	fc_delKO
1	1.235	1.419	1.226

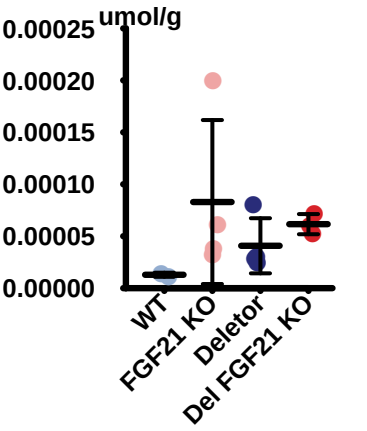
5,10−Metheny−THF



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO
0.852	0.1008	0.6585	0.0711	0.4923	0.1419

fc_wt	fc_KO	fc_del	fc_delKO
1	1.123	1.039	1.495

5,10−Methylene−THF



p_Del_wt	p_Del_DelKO	p_Del_KO	p_DelKO_wt	p_KO_wt	p_DelKO_KO

fc_wt	fc_KO	fc_del	fc_delKO
1	6.415	3.161	4.769